Structural Stability of Nucleic Acids and Peptides: a Theoretical and Computational Study

Author: Zuojun Guo

Boston College

The Graduate School of Arts and Science

Department of Chemistry

STRUCTURAL STABILITY OF NUCLEIC ACIDS AND PEPTIDES: A THEORETICAL AND COMPUTATIONAL STUDY

dissertation

by

ZUOJUN GUO

submitted in partial fulfillment of the requirement

for the degree of

Doctor of Philosophy

May 2012

STRUCTURAL STABILITY OF NUCLEIC ACIDS AND PEPTIDES:

A THEORETICAL AND COMPUTATIONAL STUDY

STRUCTURAL STABILITY OF NUCLEIC ACIDS AND PEPTIDES: A THEORETICAL AND COMPUTATIONAL STUDY

Zuojun Guo

Advisor: Professor Udayan Mohanty

ABSTRACT:

In chapter one, two simple models are used to estimate the electrostatic contributions to the stiffness of short DNA fragments. The first model views DNA as two strands that are appropriately parameterized and are wrapped helically around a straight cylinder radius equal to the radius of the DNA molecule. The potential energy of the DNA due to phosphate-phosphate electrostatic interactions is evaluated assuming that the charges interact through Debye-Hückle potentials. This potential energy is compared with the potential energy as computed using our second model in which DNA is viewed as two helical strands wrapping around a curved tube whose cross-section is a disk of radius equal to the radius of the DNA. The results are compared with counterion condensation models and experimental data (Guo *et al*. *J. Phys. Chem. B*, **2008**, 112, 16163-16169).

In chapter two, the fidelity of translation selection begins with the base pairing of codon-anticodon complex between the mRNA and tRNAs. Binding of cognate and near-cognate tRNAs induces 30S subunit of the ribosome to wrap around the ternary complex, EF-Tu(GTP)aa-tRNA. We have proposed that large thermal fluctuations play a crucial role in the selection process. The binding energies of

over a dozen unique site-bound magnesium structural motifs are investigated and provide insights into the nature of interaction of divalent metal ions with the ribosome (Guo *et al*. *Proc. Nat. Acad. Sci.* **2011**, 108, 3947-3951).

In chapter three, we use extensive molecular dynamics simulations to study a series of stapled alpha helical peptides over a range of temperatures in solution. The peptides are found to exhibit substantial variations in predicted helicities that are in good agreement with the experimental value. In addition, we find significant variation in local structural flexibility of the peptides with the position of the linker, which appears to be more closely related to the observed differences in activity than the absolute α-helical stability (Guo *et al*. *Chem. Biol. Drug. Des*. **2010**, 75, 348-359.).

In chapter four, the α-helical conformation and structural stability of single and double stapled all-hydrocarbon cross-linked p53 peptides in solution and when bound to MDM2 is investigated. We determined the effects of the peptide sequence, the stereochemistry of the cross-linker, the conformation of the double bond in the alkene bridge, the length of the bridge, on the relative stability of the α-helix structure. The conformation population distribution indicates a fully helical state and several partially folded states. The distribution of dihedral pairs of the stapled peptides in the bound state indicates a significant population around the α-helical region. Sequences over which the linker spans tend to have the highest helical occupancy. Significant helical content is observed for a double stapled p53 peptide at 575 K. The probability to form native contacts is increased

when the stapled peptides are bound to MDM2. The distribution of the end-to-end distance of the peptides is bimodal.

Table of Contents

Glossary of Terms

Chapter 1

Counterion condensation. The phenomenon in which counterions from aqueous solution condense on the backbone of polyelectrolyte until its linear charge density is reduced below a threshold value.

Bjerrum length. It is the length at which the electrostatic interaction between two elementary charges is comparable in magnitude to the thermal energy.

Worm like chain model (WLC). The WLC model of a polymer chain assumes that it behaves as an isotropic flexible rod. In contrast, a freely-jointed chain is flexible only between discrete segments.

Persistence length. If one takes a segment of a polymer chain that is shorter than the persistence length, the molecule behaves like a flexible elastic rod. For segments of the chain longer than the persistence length, its conformation properties of the chain resemble that of a three-dimensional random walk.

Debye screening length. A Debye sphere is a volume around a given charge outside of which charges are screened by the salt ions. The radius of the sphere is called the Debye screening length; this quantity depends on the square root of the ionic strength.

J-factor. It is a measure of the relative intra-molecular end-to-end concentration. This quantity is related to chain flexibility.

Debye Hückle potential. The potential energy of interaction between a pair of charges that is screened by other charges in aqeuous solution.

Linear charge density. It is defined as the ratio of the Bjerrum length to the spacing of the charges along the backbone of the polymer chain. For B form DNA, the linear charge density is 4.22.

Chapter 2

Translation. In translation the messenger RNA (mRNA) is decoded by the ribosome to produce specific amino acid chain that will fold into an active protein.

Ribosome. It is a component of cell that assembles the twenty specific amino acid molecules to form a particular protein molecule.

Peptidyl transferase region (PTC). PTC resides in the large subunit of the ribosome. PTC catalyzes the formation of the peptide bond as well as the release of the peptide.

Single-molecule fluorescence energy resonance transfer (smFRET). A single molecule biophysical technique that can measure distances at the 1-10 nanometer scale length and time scales in the 10-100 ms. A donor and acceptor FRET pair that are appropriately placed in the molecule are excited and detected by lasers at appropriate wavelengths. In smFRET, heterogeneous processes can be studied without ensemble averaging.

Gaussian Markov processes. In Brownian motion theory, the noise is Gaussian if its correlation function is a normal distribution. The noise is a Markov process if its two-time correlation function is a delta function.

Binding energy. It is the energy to disassemble a molecule into its separate parts.

QM/MM. This is a simulation method that is based on partitioning a biochemical system into two parts: an electronically important region (QM) and the remainder (MM). The QM region is described by high-level *ab initio* or density functional quantum mechanical approximations. The remaining region is a perturbation and is described by classical approximation, such as that based on molecular mechanics (MM) force fields.

Chapter 3

Tumor suppressor gene p53. It is a tumor suppressor protein that is encoded in humans by the *TP53* gene. In multicellular organisms, p53 regulates the cell cycle and functions as a tumor suppressor.

Transcriptional factor. It is a protein that binds to specific DNA sequences. This protein controls the flow of genetic information from DNA to mRNA.

E3 ubiquitin ligase. It is a protein that in conjunction with ubiquitin enzyme allows the attachment of ubiquitin to a lysine on another protein.

MDM2. It is the name of a gene and the protein encoded by it. MDM2 functions as E3 ubiquitin ligase. It also inhibits p53 transcriptional activation.

Stapled α-helical peptide. This is a strategy for stabilizing α-helices through an all-hydrocarbon cross-link. The key components of this strategy are α-methylated amino acids which bear olefinic side-chains of varying length and configured with

either *R* or *S* chirality. These amino acids are incorporated into the peptide at the *i* and either *i*+4 or *i*+7 position, and then connected through olefin metathesis to cross-link one or two turns of the helix.

Desmond. It is a software package (Schrodinger Inc.) that allows one to carry out robust molecular dynamics simulations of biological systems.

Replica exchange molecular dynamics (REMD). *N* identical copies, each at different temperature, of the system are subjected to molecular dynamics simulations. Using Metropolis criterion, one exchanges configurations between the replicas. This allows configurations at high temperatures available to the simulations at low temperatures and vice versa.

Circular dichroism (CD). It is a spectroscopic tool that measures the difference in the absorption spectra of right-handed and left-handed circularly polarized light when it impinges on matter that has structural asymmetry. In proteins, for example, beta sheets and alpha helices are chiral.

Potential energy landscape of proteins. The potential energy of the various conformation states of a protein during its motion that includes the bad contacts between residues, as a function of reaction coordinates, the later could be the solvent-averaged energy and the fraction of native contacts. It is postulated that folding of a protein occurs through an ensemble of structures in which the potential energy landscape is funnel shaped with multitude of traps. The lowest energy folded structure is the native state of the protein.

Chapter 4

Native contact. It is those contacts between the side chains of two amino acids residues that are not neighbors in the sequence space but are spatially close in the protein structure. The fraction of native contacts is used as a reaction coordinate in analyzing molecular dynamics simulations of proteins.

Contact map. The map describes the distances between residue pairs in a protein using a two-dimensional matrix. The *ij* element of the matrix is unity if residues *i* and *j* are less than a predetermined value, and is zero otherwise.

End-to-end distance. Consider a vector that points from one end to the other end of a polymer chain. The magnitude of the vector is the end-to-end distance.

Ring-closing olefin metathesis (RCM). RCM is an olefin metathesis reaction that allows closing of 7-8 member rings.

Introduction

In this thesis, I have investigated by theoretical and computational methods the following questions: What is the persistence length of a tightly bent DNA? How does the persistence length of DNA depend on monovalent salt concentration? What is the role of large thermal fluctuations of the ribosome in driving the ternary complex towards the GTPase activated state? What are the binding affinities of the magnesium binding sites observed in the various crystal structures of the ribosome? What is the relationship between the structure and the dynamic stability of hydrocarbon linked stapled p53 peptides? How is the helical configuration of various stapled peptides in solution modified by the characteristics of its conformation preference when bound to MDM2?

I will now summarize some of the basic ideas behind the various topics investigated: (A) the electrostatic contribution to DNA stiffness; (B) the role of large thermal fluctuations and magnesium ions in tRNA selectivity of the ribosome, and (C) the conformation and structural stability of all hydrocarbon *cis* and *trans* linkers single and double stapled to p53 peptides that are bound to MDM2.

A. DNA electrostatics

DNA is a locally stiff charged polymer that displays resistance to bending and twisting.[1-15] The stiffness of DNA is nearly constant at ~450 - 500 Å in monovalent salt concentration between 0.01 M and 1.0 M, while its twisting

modulus is between 2.4 x 10^{-19} and 4.5 x 10^{-19} erg.cm. In cells, the rigidity of DNA is modified by variety of proteins that allows it to fold and loop.[1-7]

The nature of electrostatic interactions in highly charged biopolymers like DNA is intricate due to the phenomenon of counterion condensation discovered by Onsager and Manning.[8,9] The basic idea can be summarized as follows. The closely placed phosphate charges along the DNA backbone leads to strong electrostatic repulsion. However, when the DNA is immersed in an aqueous solution containing monovalent or divalent cations from added salts, the thermodynamic property of system is electrostatically unstable.[8,9] The free energy of the system is lowered when counterions from the bulk condense on the backbone of the DNA. Manning and Onsager showed that as a result of counterion condensation, each phosphate charge is reduced by the factor 1-θ, where θ is the number counterions associated per phosphate charge. Furthermore, θ does not vary with salt concentration up to about 0.1 M and depends only on the valence of the counterion, the spacing between the phosphate charges, and the Bjerrum length -- the characteristic distance at which coulomb repulsion between a pair of charges balances thermal fluctuations. [8,9]

Counterion condensation modifies the electrostatic contribution to the free energy associated with assembling the DNA from the phosphate charges that were initially far apart in aqueous solution. Consider the free energy of a straight DNA. If the electrostatic interaction between a pair of phosphate charge is

governed by Debye-Huckel potential, then the free energy of assembling a straight DNA, i.e., DNA whose conformation is as depicted in Figure 1, is[8,9]

$$\frac{G_1}{k_B T} = \frac{2q_{net}^2}{\varepsilon k_B T b} \sum_{i=1, i\neq j}^{N} \sum_{j=1}^{N} \frac{e^{-\kappa|i-j|b}}{|i-j|} \tag{1}$$

where k_B is the Boltzmann constant, N is the number of phosphate charges, T is the absolute temperature, q_{net} is a renormalized phosphate charge is expressable in terms of the bare phosphate charge q as q(1-zθ), z is the valence of counterion, ε is the bulk dielectric constant of the solvent, and as already stated, θ is the number counterions associated per phosphate charge. The phosphate charges are labeled by indices *i* and *j*, and *κ* is the inverse of the Debye screening length and is related to the ionic strength.[8,9]

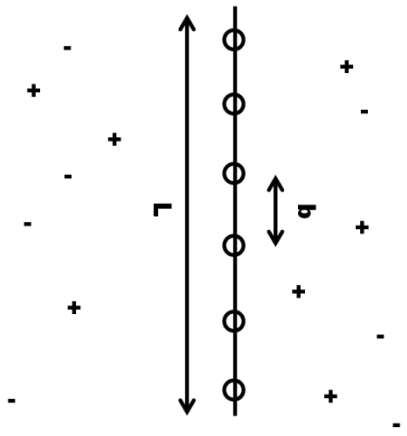


Figure 1. A line model of DNA is shown. b is the distance between two phosphate charges and L is the contour length of the DNA. The counterions and the anions are indicated by - and + respectively.

The double sum in equation (1) can be simplified to yield[8,9]

$$\frac{G_1}{k_B T} = -\xi(1 - z\theta)^2 \ln(1 - e^{-\kappa b}) \tag{2}$$

where ξ is the linear charge density of DNA, and is defined as the ratio of the Bjerrum length l_b to the charge spacing b. For B-DNA, ξ is 4.2 at room temperature.

According to Manning-Onsager hypothesis, the counterions from the solvent will condense in a small volume around the DNA.[8-11] Consequently, there is an entropic contribution to the free energy, G_2, arising from bringing the counterions from the bulk solution to a small region in the vicinity of the DNA.[8-11]

$$\frac{G_2}{k_B T} = \theta \ln\left(\frac{10^3 \theta}{c\gamma z Q}\right) \tag{3}$$

In the above relation, *c* is the salt concentration, *γ* is the activity coefficient of the solvent, and *Q* is an effective volume within which the condensed counterions reside. The factor 10^3 is introduced so that the unit of *Q* is in cm^3/mol.

Since the system is not at equilibrium, we must minimize the total free-energy with respect to θ. [8-11]

$$\begin{aligned} &\frac{\partial}{\partial\theta}\left[\frac{G_1 + G_2}{k_B T}\right] = 2z(1 - z\theta)\xi \ln\left(\frac{1 - e^{-\kappa b}}{\kappa b}\right) \\ &+2z(1 - z\theta)\xi\left[\ln\left[b\left(8\pi 10^{-3} N_A l_b\right)^{1/2}\right] + \frac{1}{2}\ln c\right] + \ln\left(10^3 \theta c\gamma z Q\right) + 1 = 0 \end{aligned} \tag{4}$$

Under limiting law conditions, i.e., as the salt concentration *c* goes to zero, we obtain an explicit expression for *Q* and θ[8-11]

$$z\theta = 1 - \frac{1}{z\xi} \tag{5}$$

$$Q = \frac{8\pi e N_A b^3}{\gamma z^4}\left(\xi - \frac{1}{z}\right)\left(\frac{1-e^{\kappa b}}{\kappa b}\right)^2 \tag{6}$$

On substituting the expressions for $z\theta$ and Q in G, we obtain the minimized electrostatic free energy of the DNA. [8-11]

$$\frac{G}{k_B T} = \frac{1}{z}\left(\frac{1}{z\xi} - 2\right)\ln\left(1-e^{\kappa b}\right) - \frac{1}{z} + \frac{1}{z^2\xi} \tag{7}$$

DNA bending and persistence length

In this section we consider the electrostatics of a weakly bent model of DNA as shown in Figure 2. The distance between the phosphate charges r_n as a function of circular radius R is[8-11]

$$r_n = \sqrt{2}R\left[1-\cos\frac{nb}{R}\right]^{1/2} \xrightarrow{\text{Taylor Series}} r_n = nb\left[1-\frac{1}{24}\left(\frac{nb}{R}\right)^2\right] \tag{8}$$

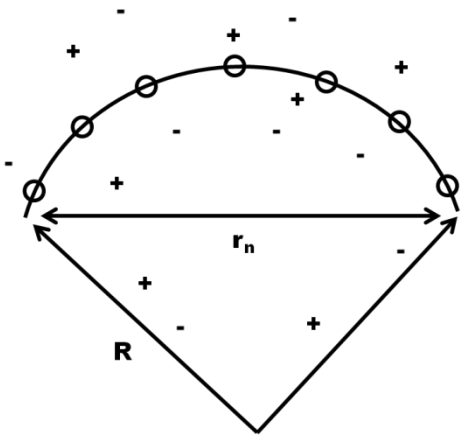


Figure 2. Weakly bent line model of DNA is shown. R is the radius of curvature.

Using the Debye-Hückel potential to describe the electrostatic interactions, one obtains the electrostatic free energy of assembling the DNA in the bent conformation[8-11]

$$\begin{aligned}\frac{G_1(R)}{k_BT} &= \frac{2q_{net}^2}{\varepsilon k_BT}\sum_{n=1}^{\infty}\frac{e^{-\kappa r_n}}{r_n} = \frac{2q_{net}^2}{\varepsilon k_BT}\sum_{n=1}^{\infty}\frac{e^{-\kappa nb}}{nb}e^{-\frac{\kappa(nb)^3}{24R^2}}\left(1-\frac{(nb)^2}{24R^2}\right)^{-1} \\ &\approx \frac{2q_{net}^2}{\varepsilon k_BT}\sum_{n=1}^{\infty}\frac{e^{-\kappa nb}}{nb}\left(1+\frac{(nb)^2}{24R^2}\right)\left(1+\frac{\kappa(nb)^3}{24R^2}\right) \\ &\approx \frac{2q_{net}^2}{\varepsilon k_BT}\left(-\frac{\ln(1-e^{-\kappa b})}{b}+\frac{b}{24R^2}\frac{e^{-\kappa b}}{\left(1-e^{-\kappa b}\right)^3}\left[\kappa b-1+e^{\kappa b}\left(1+\kappa b\right)\right]\right)\end{aligned} \tag{9}$$

We assume that the entropic term of the free energy will not vary too much with DNA curvature for weak bending[8-11]

$$\frac{G_2(R)}{k_BT} = \theta\ln\left(\frac{10^3\theta}{c\gamma zQ}\right) \tag{10}$$

Minimizing the total free energy with respect to θ yields[8-11]

$$\begin{aligned}&\frac{\partial}{\partial\theta}\left[\frac{G_1(R)+G_2(R)}{k_BT}\right] = 2z(1-z\theta)\xi\ln\left(\frac{1-e^{-\kappa b}}{\kappa b}\right) \\ &+2z(1-z\theta)\xi\ln(\kappa b)\ln\left(10^3\theta c\gamma zQ\right)+1 \\ &-2z(1-2z)l_b\frac{b}{R^2}f(\kappa b)=0\end{aligned} \tag{11}$$

where the quantity f(x) is defined as

$$f(\kappa b) = \frac{e^{-\kappa b}}{24\left(1-e^{-\kappa b}\right)^3}\left[1+\kappa b-e^{-\kappa b}\left(1-\kappa b\right)\right] \tag{12}$$

In the limit of low salt concentrations, we get

$$z\theta = 1 - \frac{1}{z\xi} \tag{13}$$

$$Q = \frac{8\pi e N_A b^3}{\gamma z^4}\left(\xi - \frac{1}{z}\right)\left(\frac{1-e^{\kappa b}}{\kappa b}\right)^2 e^{-\frac{2b^2}{R^2}} f(\kappa b) \tag{14}$$

On substituting these expressions in G we obtain the total free energy of the weakly bent DNA.[8-11]

$$\frac{G(R)}{k_B T} = \frac{1}{z}\left(\frac{1}{z\xi} - 2\right)\ln\left(1-e^{\kappa b}\right) - \frac{1}{z} + \frac{1}{z^2\xi} + 2\theta\frac{b^2}{R^2} f(\kappa b) \tag{15}$$

According to the worm like chain model,[12,13] the electrostatic persistent length P_e is obtained from the difference of free energy between a weakly bent DNA and a straight DNA of identical length,

$$\Delta G = \frac{L k_B T}{2R^2} P_e \tag{16}$$

In the line model, the difference between the free energy between a weakly bent DNA and a straight DNA of identical length is[8-11]

$$\frac{G_1(R)}{k_B T} = \frac{G_1}{k_B T} + (1 - z\theta)^2 l_b\left(\frac{b}{R^2}\right) f(\kappa b) \tag{17}$$

Thus, the electrostatic contribution to persistence length is[8-11]

$$P_e = \frac{4b}{z} f(\kappa b)\left(1 - \frac{1}{z\xi}\right) \tag{18}$$

Observe that the persistence length dependents explicitly on the salt concentration through the Debye screening length.[13,14,15] The results are in agreement with the classic work on DNA bending by Skolinck-Fixman and Odijk.[13-15]

B. Magnesium ions and t-RNA Selectivity of the Ribosome

A wealth of experimental studies have led to our current understanding on how ribosome catalyzes various biochemical processes such as translation, and translocation.[16,17,18] In translation, the ribosome selects the correct transfer RNA (tRNA). The initial selection of aa-tRNA (aminoacyl-tRNA) consists of binding of the ternary complex (aa-tRNA, guanosine triphosphate(GTP), and elongation factor EF-Tu) on the ribosome, interaction between tRNA and mRNA (the codon recognition), and stabilization of the ternary complex at the decoding site.[17,20] During translation, ribosomal subunits associate during initial selection and dissociate during recycling after termination. The association and the various interactions that result from the movement of the ribosomal subunits are specific and dynamic.[17,20]

In the initial selection of aa-tRNA, the ternary complex binding induces structural changes in the ribosome that allows it to form stabilizing contacts after the codon anticodon recognition.[17,20] If one base-mismatch occurs out of the 3 bp (near cognate tRNA) during the codon recognition at the decoding site of the ribosome, then the stabilizing contacts is weaker compared with completely match base pairing (cognate tRNA) 3-nt codon of the mRNA with the anticodon end of tRNA.[17,20] Base pairing of mRNA with tRNA occurs at a frequency of 20-30 per second. The differences in the free energy associated with base-pairing of mRNA codon with tRNA anticodon cannot explain the low error rate of 10^{-4}. [17,20]

Role of magnesium ions

Metal ions such as sodium and magnesium serve important biological functions through adopting and stabilizing proper tertiary structure of RNA, such as the genomic organization and cellular metabolism, including protein transport, RNA splicing, peptide bond formation, and translational regulation.[18,19] Magnesium is one of the most abundant intracellular multivalent cations. It has the highest charge density of all biologically relevant cations due to its relatively small ionic radius.[18] As a result, magnesium ions can effectively neutralize the backbone phosphates and stabilize the conformation of RNA and DNA.

It is generally believed that magnesium ions bind to RNA either by diffuse or site binding.[19-21] In diffuse binding, RNA interacts with hydrated magnesium ion by long-range electrostatic interaction.[19-21] An example of diffuse binding is counterion condensation of magnesium ions. These divalent ions form a delocalized cloud around RNA and substantially neutralize the negative backbone phosphate charges and provide thermodynamic stability.[20] In site binding anionic ligands in RNA and partially dehydrated magnesium ions interact through short range interactions.[18-21]

The fidelity of protein synthesis is sensitively dependent on the presence of magnesium ions.[17-21] In fact, high resolution crystallographic structures of *H. marismortuni*, *E. coli*, and *T. thermophilus* ribosomal subunits reveal that approximately 40% of the magnesium binding sites are conserved across the kingdoms of life.[18-20, 22] Furthermore, ribosomal subunits can reversibly dissociate

and associate in vitro as a function of Mg concentration;[23] this suggests that the stability of the intersubunit contacts are governed by divalent ions.[24]

The crystal structure of *H. morismortuni* ribosome reveals over 116 magnesium binding sites in the large subunit of the ribosome.[18] Magnesium ions are observed to have zero (type 0), one (type I), two (type II), three (type III), four (type IV), five (type V) inner-sphere contact with RNA or protein ligands.[18] These sites have been cataloged into six different types depending on the number of direct interactions (inner-sphere contact) of the divalent ion within each motif.[18] According to the geometric arrangement of these contacting bonds, these magnesium site also can be divided into some subgroups, such as type IIa, type IIb, type IIIa, type IIIb (Figure 3). Magnesium ions of type IIa are the most common type with 40 sites observed (34.5%). Comparing with type IIa, type IIb are much less obtained, only 5 examples (4.3%) have been observed. Mg^{2+} ions of type I is the second common type with 37 examples (31.9%), Mg^{2+} ions of types IIIa, IIIb, IV and V possess 17.2% of the total, and there are five Mg^{2+} ions which are difficult to locate the coordinated waters, all of them are grouped into type X.[18] The observed frequencies of each type of Mg^{2+} ions in 23S subunit can be used as reflection of how important the roles Mg^{2+} ions played in the stabilization of this ribosomal subunit structures.[18]

Contacts between Mg^{2+} ions and ribosomal proteins are more often made through the outer sphere (32 contacts) than through the inner sphere (13 contacts) of Mg^{2+} ions.[18] The contacting mode between Mg^{2+} ions and the

ribosomal atoms also can be categorized as inner-sphere contact and outer-sphere contact. The contacts between nonbridging phosphate oxygen atoms of the RNA backbone with Mg^{2+} ions account for 73% of the inner-sphere contacts; contacts between nonbridging phosphate oxygen atoms of the RNA backbone with Mg^{2+} ions are about 50% of the outer-sphere contacts.[18] Concentration of Mg^{2+} ions is highest in regions of 23S RNA that are most conserved across the three domains of life, microchondria and chloropasts.[18-20] Approximately half of Mg^{2+} ions interact with peptidyl transferase region (PTC) and the conserved regions of domain II and IV adjacent to PTC. [18-20]

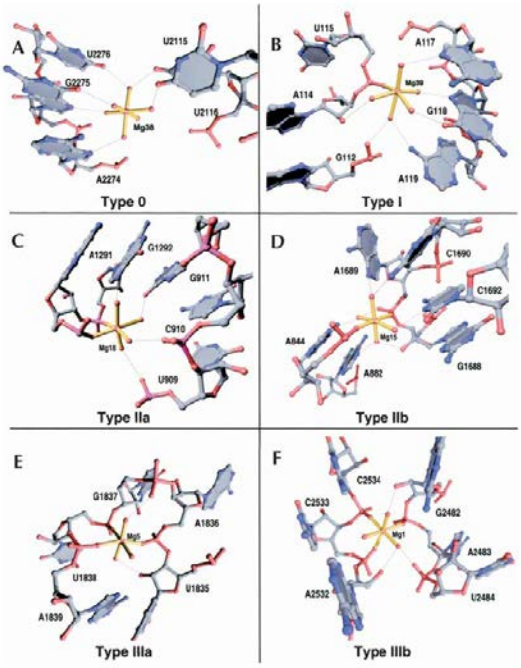


Figure 3. Example of geometric classes of Mg^{2+} ions in the large ribosomal subunit. Figure is adopted from reference 18.

In their seminal work on the pretranslocation state within the 70S of the *T. thermophilus* ribosome, Selmer *et al* argue that the magnesium ions aid intersubunit interactions and association.[24,20] The crystal structures reveal that the interactions between ribosomal subunits occur through various bridges that are critical for the association of the subunits. These bridges have unique characteristics. A bridge can change conformation when tRNA binds to it.[24,20] There is a bridge in the ribosome that is entirely mediated by a divalent metal cation.[24,20] In contrast, another bridge has been identified in which both divalent metal ions and protein side chains contributed to the bridge formation and stability.[24,20]

In bridge B2 in *E.coli*, the ribose O4' of A1913 of H69 and a phosphate oxygen of position 38 of A-site tRNA is bridged by a Mg^{2+} ion, while the ribose 2'-OH and the nonbridging phosphate oxygens of 16S 1493 and 1494 is bridged by the a different Mg^{2+} ion.[20,24] In B2c, Mg^{2+} ions mediate interactions between the backbones of h24 and h27 of 16S RNA and H67 of 23S RNA. In bridge B5, there are interactions between Arg^{49} of L14 and a non-bridging phosphate oxygen of nucleotide 1423 of 16S RNA.[20,24] The interesting point of this bridge is that one Mg^{2+} ion is coordinated to the non-bridging phosphate oxygens of 1421 of h44, 1950 and 1951 of H71, and Glu^{54} of L14.[20,24]

Single molecule (smFRET) studies of the ribosome

The development of single-molecule fluorescence energy resonance transfer (smFRET) methods has provided us with a fundamental tool to investigate the

structure, the dynamics, and the function of the ribosome.[20] Chu and coworkers have carried out extensive smFRET investigation of initial selection, translation, and translocation of aa-tRNA in *E. coli* ribosome.[17,20] In a typical smFRET experiment, the ribosome is preassembled with biotinylated mRNA and Cy3-labeled peptidyl-tRNA (P-site tRNA).[17,20]

$tRNA^{fMet}$ was labeled with Cy3 at the 4-thiouriding (s^4U) residue present at position 8, and $tRNA^{Phe}$ was labeled with Cy5 at the 3-(3-amino-3-carboxypropyl) uridine (acp^3U) residue at position 47.[17] These labeling will not affect the function of tRNA since position 8 and 47 are located at the elbow region of tRNA. During these smFRET experiments, Cy3 is illuminated, and Cy3 and Cy5 emission intensities (I) are simultaneously monitored with the FRET value defined as $I_{Cy5}/(I_{Cy3}+ I_{Cy5})$.[17] When smFRET ≥ 0.25, it indicates that the distance between molecules labeled with Cy3 and Cy5 is within 60 Å. Three peaks with mean FRET values centered at 0, 0.45, and 0.74 is observed in the time-resolved population FRET histogram.[17] The low 0 FRET state are observed when peptidyl-tRNA dissociated from the A site. The high 0.74 FRET state indicates the closest approach between the two tRNA molecules.

The smFRET studies of the pathways in tRNA selection process in *E. coli* ribosome indicated that base pairing of codon-anticodon complex involve not only interactions with specific residues and ribosomal proteins, but also conformation fluctuations of the ribosome.[17,20] These conformation changes cause the small subunit to fold around the ternary complex.[17,20] As a result, the

cognate ternary complex rotates into an orientation that allows it to make stabilizing contacts with functional groups such as GAC and the sarcin-ricin loop (SRL) near the GTPase activated state.[17,20] The time-scale for forming the stable contacts is between 65 ms and 100 ms at 15 mM and 5 mM Mg^{2+} concentrations respectively, while the distance traversed by the ternary complex in going to GTPase activated state is over 70 Å.[17,20] Consequently, large thermal fluctuations of the ribosome are essential to rotate the ternary complex into a position so as to facilitate stable contacts with GAC and the sarcin-ricin loop in the large subunit.[17,20] According to these experiments, large and rate thermal fluctuations play a critical role in the low error rate selectivity process. [17,20]

C. p53 peptide

The tumor suppressor gene p53 is responsible for maintaining the integrity of the human genome and plays a vital role in DNA repairing machinery.[25] p53 consists of two folded domains: the DNA-binding core domain (CD; residues 94-294) and the tetramerization domain (TetD; residues 323-360). It also has two disordered domains: N-terminal domain (NTD; residue 1-94) and C-terminal domain (CTD; residues 360-393). The disorder domains mediate the interactions with proteins such as p300/CBP, MDM2, and the S100 family.[25]

MDM2 is a 491 amino acid residue protein that has a N-terminal p53-binding domain, a central domain, a zinc finger domain, and a C-terminal zinc-dependent RING finger domain. In contrast, MDMX has 490 amino acid residues and

possesses domain structures that are arranged similar to MDM2, except that it lacks ubiquitin-ligase function.[26]

Loss of p53 tumor suppressor activity is a frequent defect in ≈50% of human cancers.[27] Defective p53 contributes to tumor growth and malignant progression.[26] The defective p53 protein is due to mutations, deletions, or overexpression of hDM2 (human homolog of MDM2).[28] MDM2 controls the stability of p53 through ubiquitination to target the tumor suppressor protein for degradation by the proteasome.[26] Inhibition and disruption the interactions between p53 and the E3 ubiquitin ligase MDM2/MDMX will reactivate the p53 pathway and selectively kill tumor cells.[26]

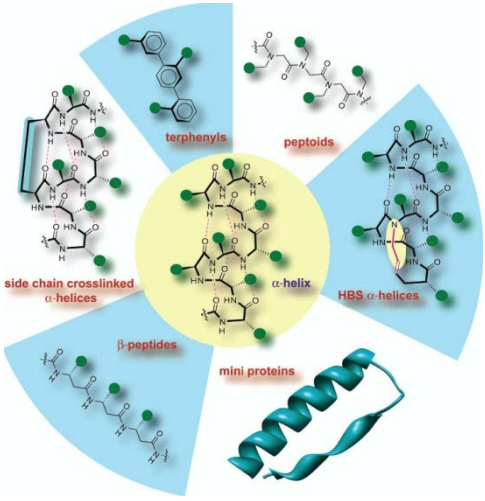
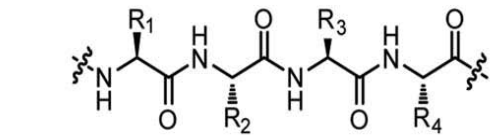


Figure 4. Strategies that stabilize the α-helical structures in peptides or mimics that include side chain crosslinking system, terphenyl helix mimetics, peptoids, HBS α-helices, mini proteins, and β-peptides. Figure is adopted from reference 29

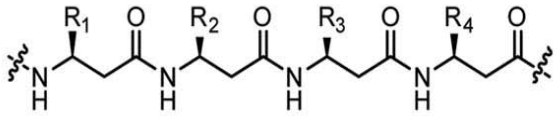
The quaternary structure of human p53 has been solved by a combination of NMR spectroscopy, electron microscopy, small-angle X-ray scattering (SAXS), single molecule FRET (smFRET), and time resolved FRET (trFRET).[25] The unique advantage of smFRET over the usual ensemble averaged techniques is that it allows us to quantify the heterogeneity of the ensemble. This is important in proteins that are conformationally mobile, such as in the denatured state.[25]

α-helical conformation of peptides

Most peptides that bind to their receptors in α-helical conformation have little helical structure in free solution.[30] Stabilizing the helical structure would lead to high binding affinity with receptors.[30] Several strategies have been developed to stabilize the α-helical conformation in peptides or mimics with non-natural scaffolds. These approaches can be divided into three categories: helical structure stabilization, which includes side-chain crosslinked peptide, miniproteins, and hydrogen bond surrogate; helical foldamer which includes β-peptide helices and peptoid helices; and helical surface mimetics which include terphenyl helical mimetics.[31]



α-peptide



β3-peptide

Figure 5. Comparison of the structures of α-peptide and β^3-peptide. Figure is adopted from reference 31.

A difference between β-peptide and α-peptide is an additional backbone carbon atom (Figure 5).[32] β-peptides can fold into helices, sheets, and turns. The helical conformations of a β-peptide are relatively stable in aqueous solution.[32] β-peptides bind to proteins such as hDM2 and hDMX and inhibit their interactions with α-helical ligands in vitro. β-peptides are usually not cell-permeable.[32] Recently, a highly cationic β^3-peptide has been synthesized that is cell permeable to upregulate p53-dependent genes in live cells.[33]

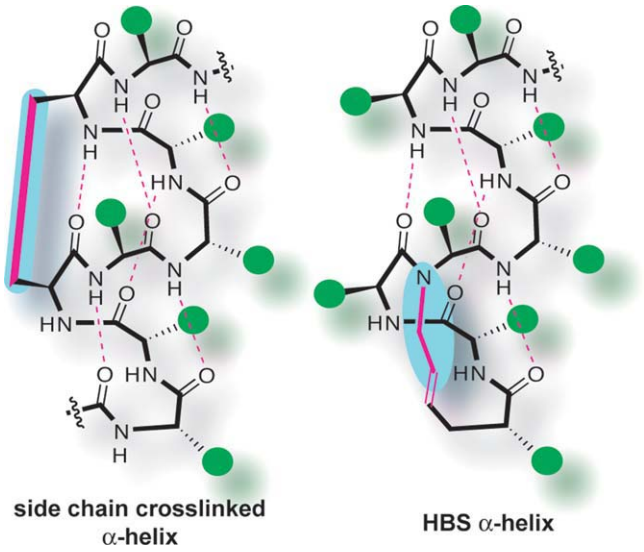


Figure 6. Comparison of the structures between side chain cross-linked α-helix and hydrogen bond surrogate α-helix. Figure is adopted from reference 31.

The hydrogen bond surrogate (HBS) approach was first investigated by Arora et al.[31] In this approach one replaces one of the main chain intramolecular hydrogen bonds with a covalent linkage (Figure 6).

The identification of synthetically 'stapled' α-helical peptides with cell-penetrating properties has received considerable interest due to possible therapeutic targets.[34] Examples of stapled peptides include p53,[28] BID BH3,[35,36] BAD BH3,[37] NOTCH,[38] and HIV-1 capsid.[39] Both single-turn and double-turn stapling chemistry has been developed to staple these peptides.[28,35-39] However, the relationship between the structure and the dynamic stability of stapled peptides are far from well understood. Recently, an all-atom Monte Carlo folding simulation study comparing unmodified peptides derived from RNase A and BID

BH3 with varying stapling chemistry has been reported.[40] A dynamical model reproducing the structural stability of a series of BH3-stapled peptides as a function of cross-linking moiety has been investigated.[41]

In Chapters 3 and 4, we use extensive molecular dynamics simulations to study a series of stapled α-helical peptides over a range of temperatures in solution. The peptides are found to exhibit substantial variations in predicted α-helicities that are in good agreement with the experimental data.[28] In addition, we find significant variation in the local structural flexibility of the stapled peptides with the position of the linker, which appears to be more closely related to the observed differences in activity than the absolute α-helical stability.[42,43] These simulations provide new insights into the design of α-helical stapled peptides and the development of potent inhibitors of α-helical protein-protein interfaces.[42,43]

References

1. Clouter, T. E.; Widom, J. *Proc. Nat. Acad. Sci.* USA **2005**, 102, 3645.
2. Schleif, R. *Ann. Rev. Biochem.* **1992**, 61, 199.
3. Ross, E. D.; Hardwidge, P. R; Maher, L. *J. Mol. Cell. Biol.* **2001**, 21, 6798.
4. Becker, N. A.; Kahn, J. D.; Maher, J. L. *J. Mol. Biol.* **2005**, 49, 716.
5. Thomas, J. O.; Travers, A. A. *Trend in Biochemical Sciences* **2001**, 26, 167.
6. Shore, D.; Baldwin, R. L. *J. Mol. Biol.* **1983**, 170, 957-981.
7. Bryant, Z.; Stone, M. D.; Gore, J.; Smith, S. B.; Cozzarelli, N. R.; Bustamante, C. *Nature* **2003**, 424, 338-341.
8. Fenley, M. O.; Manning, G.S.; Olson, W. K. *J. Phys. Chem.* **1992**, 96, 3963-3969.
9. Manning, G.S. *Biophysical Chemistry.* **2002**,101-102, 461-473.
10. Manning, G.S. *Quarterly Review of Biophysics* **1978**, 111, 179-246.
11. Manning, G. S.; Mohanty U. *Physica A.* **1997**, 247, 196-204.
12. Schellman, J. A.; Harvey, S.C. *Biophysical Chemistry* **1995**, 55, 95-114.
13. Hagerman, P. J. *Biopolymers.* **1983**, 22, 811-814.
14. Skolnick, J.; Fixman, M. *Macromolecules* **1977**, 10, 944-948,.
15. Odijk, T. *Journal of Polymer Science.* **1977**, 15, 477.
16. Cate, J.H.; Doudna, J. A. *Structure* **1996**, 4, 1221-1229.
17. (a) Blanchard, S.C.; Gonzalez, Jr R.L.; Kim, H.D.; Chu, S; Puglisi, JD *Nature Structural Biology* **2004**, 11, 1008 – 1014. (b) Lee, T.H.; Blanchard, S.C.; Kim,

H.D.; Puglisi, J.D.; Chu, S. *Proc. Natl Acad. Sci. USA* **2007**, 104, 13661–13665. (c) Gonzalez, Jr R.L.; Chu, S; Puglisi, J.D. *RNA* **2007**, 13**,** 2091–2097.

18. Klein, D.J.; Moore, P.J.; Steitz, T.A. *RNA* **2004**, 10:1366-1379.
19. Pyle, A. M. *J. Biol. Inorg. Chem.* **2002**, 7, 679-690.
20. Guo, Z.J.; Gibson, M.; Sitha, S.; Chu, S.; Mohanty, U. *Proc. Natl. Acad. Sci.* USA. **2011**, 108(10), 3947-3951.
21. Laing, L.G.; Gluick, T. C.; Draper, D. E. *J. Mol. Biol.* **1994**, 237, 577-587.
22. Schuwirth, B.S., *et al*. *Science* **2005**, 310:827-834.
23. Kimes, B. W.; Morris, D. R. *Biochemistry* **1973**, 12, 442-449.
24. Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petry, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. *Science* **2006**, 313, 1935-1942.
25. Huang, F.; Rajagopalan, S.; Settanni, G.; Marsh, R.J.; Armoogum, D.A.; Nicolaou, N.; Bain, A.J.; Lerner, E.; Ying, L.; Fersht, A. R. *Proc. Nat. Acad. Sci. USA* **2009**,106, 20758.
26. Pazgier, M.; Liu, M.; Zou, G,; Yuan, W.; Li, C.; Li, C.; Li, J.; Monbo, J.; Zelia, D.; Tarasov, S.G.; Lu, W. *Proc. Nat. Acad. Sci. USA* **2009**,106, 4665-4670.
27. Kussie, P.H.; Gorina, S.; Marechal, V.; Elenbaas, Brian.; Moreau, J.; Levine, A.J.; Pavletich, N.P. *Science* **1996**, 274: 948-952.
28. Bernal, F.; Tyler, A.F.; Korsmeyer, S.J.; Walensky, L.D.; Verdine, G.L. *J. Am. Chem. Soc* **2007**, 129, 2456-2457.
29. Patgiri, A.; Jochim, A.L.; Arora, P.S. *Acc. Chem. Res.* **2008**, 41, 1289-1300.

30. Schafmeister, C.E.; Po, J.; Verdine, G.L. J*. Am. Chem. Soc* **2000**, 122, 5891-5892.

31. Guarracino, D.A.; Bullock, B.N.; Arora, P.S. *Biopolymers* **2010**, 95, 1-7.

32. Kritzer, J.A.; Lear, J.D.; Hodsdon, M.E.; Schepartz, A. *J. Am. Chem. Soc* **2004**, 126, 9468-9469.

33. Bautista, A.D.; Appelbaum, J.S.; Craig, C.J.; Michel, J.; Schepartz, A. *J. Am. Chem. Soc*. **2009**, 132, 2904-2906.

34. Walensky, L.D.; Verdine, G.L.; *Clin Cancer Res* **2007**, 13, 7264-7270.

35. Walensky, L.D.; Kung, A.L.; Escher, I.; Malia, T.J.; Barbuto, S.; Wright, R.D.; Wagner, G.; Verdine, G.L.; Korsmeyer, S.J. *Science* **2004**, 305, 1466-1470.

36. Walensky, L.D.; Pitter, K.; Morash, J.; Oh, K.J.; Barbuto, S.; Fisher, J.; Smith, E.; Verdine, G.L.; Korsmeyer, S.*J. Mol Cell* **2006**, 24, 199-210.

37. Danial, N.N.; Walensky, L.D.; Zhang, C.Y.; Choi, C.S.; Fisher, J.; Molina, A.J.A.; Datta, S.R. *et al*. *Nat. Med.* **2008,** 14, 144-153.

38. Moellering, R.E.; Cornejo, M.; Davis, T.N.; Del Bianco, C.; Aster, J.C.; Blacklow, S.C.; Kung, A.L.; Gilliland, D.G.; Verdine, G.L.; Bradner, J.E. *Nature* **2009**, 462, 182-188.

39. Bhattacharya, S.; Zhang, H.; Debnath, A.K.; Cowburn, D. *J. Biol. Chem.* **2008**, 283, 16274-16278.

40. Kutchukian, P.S.; Yang, J.S.; Verdine, G.L.; Shakhnovich, E.I. *J. Am. Chem. Soc.* **2009**, 131, 4622-4646.

41. Hamacher, K.; Hübsch, A.; McCammon, J.A. *J. Chem. Phys.* **2006**, 124,164907(1-8).
42. Guo, Z.; Mohanty, U.; Noehre, J.; Sawyer, T. K.; Sherman, W.; Krilov, G. *Chem. Biol. Drug. Des.* **2010**, 75, 348-359.
43. Guo,Z.; Mohanty, U.; Krilov, G. (unpublished, 2011).

Chapter One

DNA on a Tube: Electrostatic Contribution to Stiffness

1.1 Introduction

Both bending and flexibility play significant roles that allow DNA to pack into nucleosomes in eukaryotes.[1-5] In prokaryotes, gene-regulatory complexes often involve short segments of DNA that are bent into loops.[6,7] Such regulatory complexes that involve looped DNA structures are necessarily limited by the bendability and the helical twist of DNA.[6-9] In eukaryotic systems, DNA looping is increasingly seen as important, being observed or predicted in a variety of gene-regulatory complexes.[7-10]

Although DNA is a highly charged polymer, locally it is rather stiff with rather high local stiffness. In fact, over a length scale of approximately 75 bp (one half of the accepted value of the persistence length), DNA is essentially rod-like.[11,12] DNA fragments of 150 bp display an average deviation in end-to-end trajectory of ~60 degrees. The high resistance to twisting is another characteristic of DNA. The torsion modulus of DNA is between 2.4×10^{-19} and 4.5×10^{-19} erg·cm.[13,14] The rigidity of DNA in cells is modified by host of proteins that allow folding and looping needed for gene expression.[6-10]

Approximately 146 bp of DNA is able to wrap almost twice around an octamer of histone proteins in eukaryotes. The conjecture[15,16] that part of the driving force for nuclesome folding is a result of asymmetric neutralization of the DNA phosphate charges was supported by results of a series of experiments by Maher and coworkers.[17-22] In this work, selected phosphate groups across the minor groove were chemically replaced by neutral methylphosphonate

residues.[17-22] Three of these modifications were inserted on complementary DNA strands (6 total) so they were located as neighbors across the minor groove.[17-22] Consequently, the electrostatic repulsion amongst the phosphate groups was interrupted. The sites neutralized were then phased with respect to A-tracts—a strategy developed to study sequence directed bends in DNA.[23-24] Gel retardation experiments were then utilized to determine the direction and the magnitude of bending due to the asymmetric charge neutralization. Asymmetric charge neutralization of duplex DNA induced collapse towards the neutral side.[17-22]

The subtle role played by decrease or increase of local inter-phosphate repulsive forces was also investigated by Hardwidge et al.[25-27] Specifically, the experiments probed how the distribution of cationic or anionic amino acids in the yeast basic zipper (bZip) binding protein GCN4 induced DNA bending.[25] DNA bending by six ammonium ions attached to flexible propyl linkers and tethered to one face of the duplex DNA was also studied.[26] If rigid linkers were used, then there was no DNA bending. These authors also studied how the insertion of an additional anionic phosphonic acid group within the major groove affected DNA bending.[27] Since two negative charges were introduced into the major groove, and the DNA responded to this charge imbalance by bending away from the hypercharged face.

Several recent studies provided considerable insights into the polyelectrolyte nature of DNA flexibility. Recently, Podesta et. al. have used scanning force

microscopy to measure the flexibility of DNA on mica surfaces coated with various concentrations of poly-L-ornithine or spermidine.[28] These authors found that DNA flexibility increased by five-fold with increasing concentration and valence of the polyamines that are used to coat the surface of mica.[28]

Cyclization assays were exploited to measure the relative intramolecular end-to-end concentration (j-factor), a quantity directly related to flexibility, for 89-bp to 105-bp DNAs with various sequences.[6, 29] A notable result, though controversial, [30] was that 89-bp DNA appeared over 200 times easier to twist than predicted by linear elastic theories. Experimental values for the j-factor for random and nucleosome positioning sequence exceed that predicted by linear elasticity theories.[6,29] However, other studies using the same approach tend to support the predictive capabilities of the worm-like chain model.[30]

It must be appreciated that cells contain activities that dramatically enhance the apparent flexibility of DNA relative to what is observed in vitro. For example, Maher and coworkers[31-33] discovered that relative helical phasing of the activator binding sites and the basal promoter did not significantly perturb *in vitro* transcription activation in HeLa cell nuclear extract. This led to the understanding that heat-resistant HMGB proteins in HeLa cell nuclear extract enhance the apparent flexibility of DNA.[31-33]

In the present theoretical work, we study the flexibility of short DNA fragments. We seek to understand and quantify the relative contributions to DNA stiffness (persistence length) of electrostatic stretching forces between backbone

phosphates vs. non-electrostatic forces including base stacking. Our model for DNA is designed so as to take into account the double helical nature of the strands and the resulting geometries for the phosphate-phosphate electrostatic interactions. The repulsions between phosphate charges interact are calculated through the Debye Hückle potential. Our model of linear DNA has the two strands wrapped around a straight cylinder; while in the model for bent DNA, the two strands are wrapped around a tube that resembles a bicycle tire. The extent to which electrostatic effects contribute to the persistence length of short DNA fragments is determined by comparing the total electrostatic energy of the circular conformer with that of the linear conformer, and using the assumption that the phosphate charges interact via the Debye-Hückle potential. The model avoids weak bending approximation inherent in continuum elastic theory of DNA stiffness.[34-38] Predictions of our model are compared with counterion condensation models and with experimental data. Note that our model is designed so as to avoid the weak bending approximation inherent in continuum elastic theory of DNA stiffness.[34-38]

1.2 Theoretical Methods

We model linear DNA as two strands, labeled H1 and H2, that are first wrapped around a straight cylinder of radius a and length L. The center of the cylinder stretches along the x-axis from $x = 0$ to $x = L$. The helical strands are

modeled as lines drawn on the cylinder. The first helix, H1, is parameterized by $\varphi \to \vec{u}(\varphi)$, where

$$\vec{u} = (x = (b / 2\pi)\varphi, y = a\sin(\varphi), z = a\cos(\varphi)). \quad (1)$$

Thus, a change of 2π of φ takes the strand once around the circle in the (y,z) plane and b Angstroms along the x-axis. The second strand, H2, is parameterized by $\varphi \to \vec{\nu}(\varphi)$ where

$$\vec{\nu} = (x = (b / 2\pi)\varphi, y = a\sin(\varphi - \frac{12}{b}2\pi), z = a\cos(\varphi - \frac{12}{b}2\pi)) \quad (2)$$

Consequently, while strand H1 first hits the $z = a$ point on the cross sectional circle at $x = 0$, the strand H2 first hits this point when $x = 12$ Angstroms. This is approximately the separation of the minor groove.

As there are α base pairs per turn, the separation between the phosphates on either strand should be $\Delta\varphi = 2\pi / \alpha$. The coordinates of phosphate i, on strand H1, are labeled by $u_i \to \vec{u}(\varphi_i)$, where $\varphi_i = (2\pi / \alpha)(i - 1)$, and i is an integer that runs through the set $\{1,...,N\}$. Observe that the length of the cylinder is $L = (b / \alpha)N$. Since strand H2 is running in the reverse direction from H1, we place the last phosphate on H2 at the end of the cylinder, thus at $x = (b / \alpha)N$, and the first one at $x = b / \alpha$. This requires that the phosphate i on H2 has coordinates, where $\varphi_i = (2\pi / \alpha)(N + 1 - i)$.

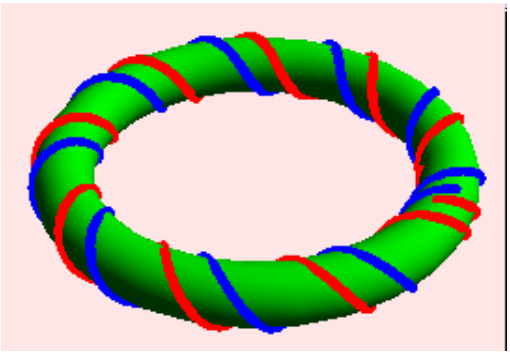


Figure 1. A double stranded DNA is wrapped around a circular tube.

Next, we model a double stranded DNA that is wrapped around a circular tube that resembles a bicycle tire (Figure 1). The tube is obtained by drawing a circle of radius a about point $(x = r + a, z = 0)$ in the (x,z) plane, and then rotating this circle around the z-axis. The resulting circle of revolution is parameterized by the pair of angles, (θ,φ), $(x = r + a + a\sin(\varphi))\cos(\theta), y = r + a + a\sin(\varphi), z = a\cos(\varphi))$. Observe that changing $\varphi \to \varphi + 2\pi$ draws a circle cross section of the tube, while $\theta \to \theta + 2\pi$ draws a circle that winds once around the z-axis, thus parallel to the central axis.

We model strand H1 on the tube as a curve that winds some number, n, times around the cross sectional circle while winding once around the z-axis. Such a curve is obtained by using the relation $\varphi = \varphi / n$. Since $\varphi \to \varphi + 2\pi n$ completes one traverse of this curve. This traverse should see N equally spaced base pairs, where the spacing is given by $\Delta\varphi = 2\pi / \alpha$. Thus, $n = N / \alpha$ implying that N must be divisible by α. If the helix is over-coiled or under-coiled relative

to α, then N need not be divisible by α. Strands H1 and H2 are parameterized respectively by $\varphi \to \vec{u}(\varphi)$ and $\varphi \to \vec{v}(\varphi)$ where

$$\vec{u}(\varphi) = (r + a + a\sin(\varphi))\cos((\alpha / N)\varphi), r + a + a\sin(\varphi)\sin((\alpha / N)\varphi), a\cos(\varphi))$$

$$\begin{aligned}\vec{v}(\varphi) = (&r + a + a\sin(\varphi - (\alpha / N)\pi)\cos((\alpha / N)\varphi), r + \\ &a + a\sin(\varphi - (\alpha / N)\pi)\sin((\alpha / N)\varphi), \\ &a\cos(\varphi - (\alpha / N)\pi))\end{aligned} \tag{3}$$

As before, the positions of the phosphates are given by taking the i[th] phosphate at the points $\varphi_i \to \vec{u}(\varphi_i)$ and $\varphi_i \to \vec{v}(\varphi_i)$ on strands H1 and H2 respectively. The parameterization of the strands assume that the minor and major grooves are compressed by the same fraction on the inside of the bicycle tube and expanded by the same fraction on the outside of the tube.

The potential energy of the double stranded DNA that is due to phosphate-phosphate electrostatic interactions is given by

$$\begin{aligned}\frac{V}{k_B T} = \frac{2q_{net}^2}{\varepsilon k_B T}[&\sum_{1 \le i \le N-1} \sum_{i < j \le N} \frac{\exp\left(-\kappa \left|\vec{u}_i - \vec{u}_j\right|\right)}{\left|\vec{u}_i - \vec{u}_j\right|} \\ &+ \sum_{1 \le i \le N-1} \sum_{i < j \le N} \frac{\exp\left(-\kappa \left|\vec{v}_i - \vec{v}_j\right|\right)}{\left|\vec{v}_i - \vec{v}_j\right|} \\ &+ \sum_{1 \le i \le N-1} \sum_{i < j \le N} \frac{\exp\left(-\kappa \left|\vec{u}_i - \vec{v}_j\right|\right)}{\left|\vec{u}_i - \vec{v}_j\right|}]\end{aligned}, \tag{4}$$

where T is the absolute temperature, k_B is the Boltzmann constant, and ε is the dielectric constant of the solvent (water).

1.3 Results and Discussion

For B-DNA, the charge spacing of the phosphates is 3.37 Å, while the linear charge density parameter ξ, defined as the ratio of the Bjerrum length to the charge spacing, is 4.22.[39]The radius a of the double helix B-DNA is taken to be 9.33 Å. The Debye screening length in monovalent salt solution is $\kappa = (\sqrt{I}/3)$Å, where I is the ionic strength in units of molarity.[39]

Let ΔV denote the difference in potential energy due to phosphate-phosphate electrostatic interactions of the double-stranded DNA that is wrapped around the circular tube and straight cylinder. In our calculation of ΔV , we assumed that each charge q is divided by the charge density parameter ξ to approximately account for counterion condensation effects.[39]

The energy to bend a segment of DNA of contour length L, assuming a restoring force obeying Hooke's law, is $\Delta V / k_B T = PL / 2R^2$, where R is the radius of curvature and P is the persistence length. Thus, the slope of $2\Delta V / k_B TL$ versus $1/R^2$, would yield the electrostatic persistence length of the DNA.[35,40] Figures 2-3 show a plot of $2\Delta V / k_B TL$ versus $1/R^2$ for salt concentrations at 5 mM, and 10 mM for the DNA tube model. Observe that $2\Delta V / k_B TL$ versus $1/R^2$ (Figure 2-3) is not linear in certain range of radius of curvature investigated. The non-linearity in the plot stems from the fact that the helix on the tube is over-coiled or under-coiled relative to α, where α is the

number of base pairs per turn; in this case, the number of phosphates, is not divisible by α.

The slope of the straight line at 5 mM salt, for DNA of lengths between 85 bp and 105 bp and between 105 bp and 130 bp, is 104.83 Å (31 bp) and 125.64 Å (37 bp), respectively. The corresponding slope at 10 mM, for DNA of lengths between 85 bp and 105 bp and between 105 bp and 130 bp, is 65.12 Å (19 bp) and 76.05 Å (23 bp), respectively. For DNA fragments between 300 and 1000 bp, the electrostatic persistence length is 119.6 Å (36 bp) at 10 mM.

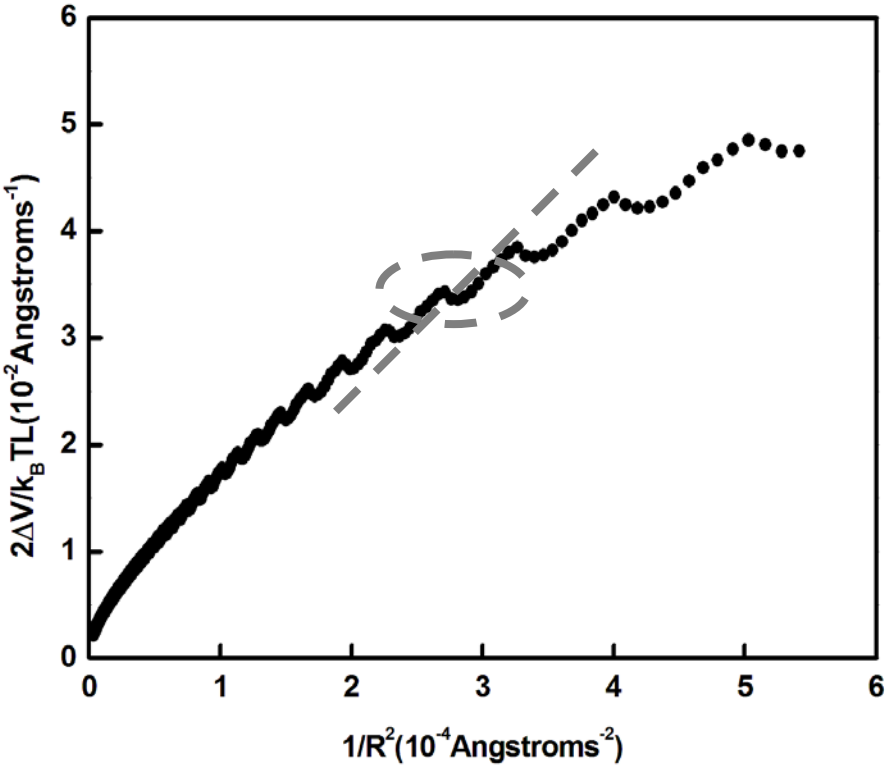


Figure 2. Plot (filled circules) of $2\Delta V / k_B TL$ versus $1/R^2$. The salt concentrate-ions is 5 mM (monovalent). The slop of the straight line through the origin (dashed line) for DNA of length between 105 and 130 bp is 125.64 Å (37 bp), which is an estimate of the electrostatic persistence length.

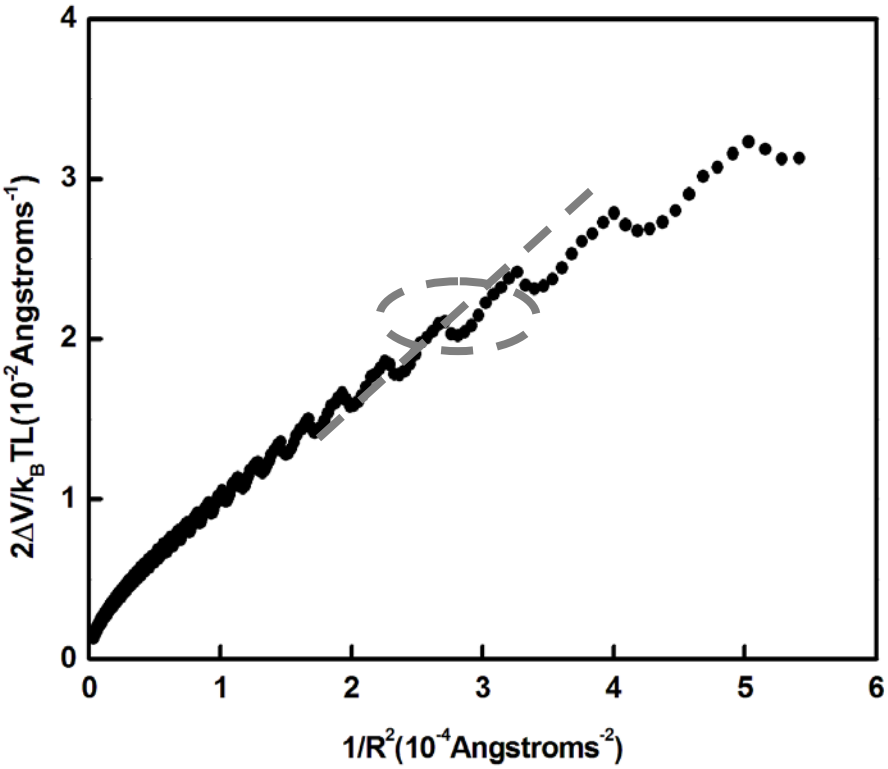


Figure 3. Plot (filled circules) of $2\Delta V / k_B TL$ versus $1/R^2$. The salt concentrations is 10 mM (monovalent). The slop of the straight line through the origin (dashed line) for DNA of length between 105 and 130 bp is 125.64 Å (37 bp), which is an estimate of the electrostatic persistence length.

Published results of ligase-mediated DNA cyclization were utilized to measure the persistence length of random DNA sequences and DNA nucleosome-positioning sequences of lengths between 89 bp to 105 bp DNAs.[6,29] In a cyclization experiment, what is measured is a quantity called the J factor; this quantity is a gauge of the flexibility of the molecules and is related th its persistence length. T4 DNA ligase-catalyzed DNA cyclization experiments are typically performed under low salt conditions and measure the apparent intramolecular end-to-end concentration (j-factor). The j-factors for the two nucleosome-positioning sequences exceed those for the three DNA random sequences.[6] Furthermore, the j-factor for all sequences examined exceeded that predicted by elastic theories of bending for all DNA constructs.[6] If the cyclization data are fitted to the entire range of DNA lengths between 89-105 bp, and for all sequences, yields a value P = 113 bp and C = 7.4 x 10^{-20} erg·cm.[6]

Inclusion of sharp kinks into the worm-like chain (WLC) bending model was shown to account for the higher-than-expected j-factors for DNA and its flexibility at short lengths (~105 bp).[30] However, such a modification leads to overestimation of the j-factors for longer DNAs where the WLC model was known to perform well.[30] Consequently, Du et. al. carried out additional measurements of j-factors for four sets of DNA fragments that were selected from phage lambda DNA to avoid sequences known to be especially curved or anisotropically flexible and which spanned the range from 105-199 bp.[30] The experimental results, for even short DNA fragments, were consistent with WLC model provided T4 DNA

ligase concentration was reduced to avoid an artifact that had affected previous experiments. The best fit to the experimental data led to helical repeat and persistence length values of 10.54 bp/turn and 470 Å (138 bp) respectively for DNA of generic sequences.[30]

When the experimental data for the J factor for random sequences is fitted to the theory for DNA cyclization based on Shimada Yamakawa linear elasticity theory of bending,[11] one obtains a value for torsional modulus C and persistence length P equal to 1.03 x 10^{-19} erg·cm and 105 bp respectively.[6] For nucleosome positioning sequences, the torsional modulus C is 7.4 x 10^{-20} erg·cm, while the persistence length P is 119 bp.[6] If the cyclization data is fitted to the entire range of DNA lengths between 89-105 bp, and for all sequences, yields a value P = 113 bp and C = 7.4 x 10^{-20} erg·cm.[6]

To compare the prediction of our model with such experimental cyclization data, we make use of the best estimate of the non-electrostatic contribution to the persistence length; this is generally regarded to be 300 Å (~89 bp).[37] However, there is a wealth of experimental evidence that the nonelectrostatic contribution to DNA stiffness in monovalent salt is between 450 and 500 Å for most, if not all, DNAs. For 0.01M NaCl, there is general consensus that P_{tot} equals or exceeds 500 Å. For a majority of DNA samples, the total persistence length is found to be nearly constant at ~450 - 500 Å between 0.01 and 1.0 M monovalent salt, and that is taken to be the intrinsic (nonelectrostatic) persistence length of DNA under those conditions.

Given this estimate, the total persistence length P predicted by our DNA tube model for short DNA fragments (85 bp to 105 bp) is 554.83 Å (165 bp) and 515.12 Å (153 bp) at 5 mM and 10 mM monovalent salt concentration, respectively (Table 1). These results suggest that electrostatic effects contribute between 18.9% and 12.6% to stiffness at these monovalent cation concentrations and DNA fragments between 85 to 105-bp.

For DNA fragments between 105 bp to 130 bp, the total persistence length obtained form the tube model is 575.64 Å (171 bp) and 526.05 Å (156 bp) at 5 mM and 10 mM monovalent salt concentration. Thus, the electrostatic effects contribute 14.5% to DNA stiffness at 10 mM for fragments between 105 to 130-bp. Overall summaries of calculations based on our DNA tube model are provided in Table 1.

Table 1. Prediction of the electrostatic persistence length of DNA based on the tube model described here at 5 mM and 10 mM (monovalent) salt solution. The best estimate of the non-electrostatic contribution to persistence length is 300 Å (89 bp).[37]

	5 mM monovalent cation			10 mM monovalent cation		
DNA Length (bp)	Electrostatic persistence length	Total persistence length	% electro-static	Electrostatic persistence length	Total persistence length	% electro-static
85-105	104.83 Å (31 bp)	554.83 Å (165 bp)	18.9%	65.12 Å (19 bp)	515.12 Å (153 bp)	12.6%
105-130	125.64 Å (37 bp)	575.64 Å (171 bp)	21.8%	76.05 Å (23 bp)	526.05 Å (156 bp)	14.5%
300-1000	356.20 Å (106 bp)	806.20 Å (239 bp)	44.2%	119.6 Å (36 bp)	569.6 Å (169 bp)	20.0%

In the DNA tube model, the electrostatic contribution to the persistence length at 150 mM (~ physiological) salt concentration for DNA fragments between 85 bp and 105 bp is approximately 5.9 Å (~ 2 bp). This result can be contrasted with the study by York *et. al*. who estimated that 30% of energy cost of deforming DNA into the curvature in the nucleosome is due to electrostatics at physiological

salt.[40] Direct comparison with this result is not possible since these authors do not compute the electrostatic contribution to persistence length. In another study, Hardwidge *et. al*. compared the electrostatic energies of two 71-bp DNA having identical sequences; one of the DNA is bent while the other is straight.[27] These authors find that the penalty in electrostatic energy for bending DNA can be attributed to crowding of phosphates which are separated over a stretch of 4-17 Å.[27] In other words, the excess electrostatics of the curved conformation is due to local phosphate-phosphate stretching forces rather than distant forces.[27]

We have compared the predictions of our model with Manning's counterion condensation models.[35,38,39] In the line model due to Manning, the difference in free energy between a straight DNA and a weakly bent DNA of identical contour length is[35]

$$\Delta G = \frac{2\theta N k_B T b^2}{R^2} f(\kappa b) \tag{5}$$

where b is the charge spacing between the phosphate charges, N is the number of phosphate charges, θ the number of counterions per charge of DNA, and $f(\kappa b)$ is defined as[35]

$$f(\kappa b) = \frac{e^{-\kappa b}}{24(1-e^{-\kappa b})^3}[1+\kappa b - e^{-\kappa b}(1-\kappa b)] \tag{6}$$

The electrostatic contribution to the persistence length based on Manning's infinite line and double helix models of B-DNA is 101.8 Å (30 bp) and 113.8 Å (34 bp) respectively at 10 mM monovalent cation concentration. The results

tabulated in Table 2 are similar to those resulting from our model (Table 1) under comparable monovalent cation concentrations.

Table 2. Prediction of the electrostatic persistence length of DNA based on counterion condensation models due to Manning[35,38,39] at 10 mM (monovalent) salt solution. The best estimate of the non-electrostatic contribution to persistence length is 300 Å (89 bp).[37]

DNA	Salt concentration	Electrostatic persistence length	Total persistence length	% electro-static
Line model	10 mM	101.8 Å (30 bp)	551.8 Å (118 bp)	18.4%
Line model	5mM	203.5 Å (60 bp)	653.5 Å (194 bp)	31.1%
Double helix model	10 mM	113.8 Å (34 bp)	563.8 Å (167 bp)	20.2%
Double helix model	5mM	221.2 Å (66 bp)	671.2 Å (199 bp)	33%

The predictions of the stiffness based on our DNA tube model, counterion condensation model[35,37-40] and others[11,34,36] assume that the nonelectrostatic and the electrostatic contributes to the total persistence length contribute in an

additive way. Recently, a different viewpoint was explored by Manning.[37] Manning proposed that the rigidity of a DNA molecule can be understood in terms of a null isomer DNA* double helix.[37] DNA* is defined to have a hypothetical structure in which all the phosphate charges have been set to zero; this is accomplished without perturbing the solvation characteristics of the molecule.[37] The statistical thermodynamics of the null isomer is constructed based on a line polyelectrolyte model and by applying a compressive force F* that is equal, but opposite, to the electrostatic tension of the charged chain.[37]

By relating the buckling persistence length in terms of F* and the bending stiffness by continuum elastic theory, an expression for persistence length of the charged DNA was obtained[37]

$$P = (r_o)^{4/3}(\pi P^* / 2)^{2/3}(F^* / k_B T) \tag{7}$$

where P^* is the persistence length of null DNA* and r_o is the radius of DNA. The compressive force $F^* / k_B T$ is a function of unsigned valence Z of counterions, the Bjerrum length, and the product κb, where b is the charge spacing and κ is the Debye screening parameter. Observe that P^* does not depend on salt concentration in contrast to other weak bending models.[11,34,36]

Using consensus value of 550 Å for the persistence length of DNA at 0.1 M, a value of 70 Å is obtained for the null persistence length P^*.[37] Furthermore, Manning showed that various experimental data[42-45] on salt dependence of the persistence length, excluding cyclization data on small DNA fragments,[6] could be

described by Eq. (7). Note that despite the fact that in the null model, P* to be only 13% of P, one cannot infer that the majority of the observed persistence length of DNA due to electrostatic effects. This is because the total persistence length in the null model cannot be obtained by adding the electrostatic and the nonelectrostatic contributions. It would be interesting to generalize our tube model to null DNA* as this would allow estimation of null persistence length P^* for sharply looped DNA.

1.4 Summary

We have proposed a model for double stranded DNA wrapped on a tube that allows quantitative estimation of the electrostatic contribution to the persistence length. Using the consensus value for the nonelectrostatic contribution (89 bp), we find that the total persistence length for a tightly bent DNA of fragment lengths between 105 to 130 bp is around 126 bp, and electrostatic effects make a contribution of 29.5% to this persistence length at 5 mM monovalent salt concentration. In contrast, this model predicts that the contribution of electrostatics to persistence length at 150 mM monovalent salt (physiological conditions) is negligible.

References

1. Travers, A. A. *Cell* **1990**, 60, 177-180.
2. Ripe, K.; Von Hippel, P. H.; Langowski, J. *Trends Biochem Sci.* **1995**, 20, 500-506.
3. Hwang, D. S.; Kornberg, A. *J. Biol. Chem.* **1992**, 267, 23083-23086.
4. Nudler, E.; Avetissova, E.; Markovstov, V.; Goldfarb, A. *Science* **1996**, 273, 211-217.
5. Gralla, J. D. *Cell* **1991**, 66, 415-418.
6. Clouter, T. E.; Widom, J. *Proc. Nat. Acad. Sci. USA* **2005**, 102, 3645.
7. Schleif, R. *Ann. Rev. Biochem*. **1992**, 61, 199.
8. Ross, E. D.; Hardwidge, P. R; Maher, L. J. *Mol. Cell. Biol*. **2001**, 21, 6798.
9. Becker, N. A.; Kahn, J. D.; Maher, J. L. *J. Mol. Biol*. **2005**, 49, 716.
10. Thomas, J. O.; Travers, A. A. *TRENDS in Biochemical Sciences* **2001**, 26, 167.
11. Shimada, J.; Yamakawa, H. *Macromolecules* **1984**, 17, 689-698.
12. Hagerman, P. J. *Ann. Rev. Biophys. Biophys. Chem*. **1988**, 17, 265-286.
13. Shore, D.; Baldwin, R. L. *J. Mol. Biol*. **1983**, 170, 957-981.
14. Bryant, Z.; Stone, M. D.; Gore, J.; Smith, S. B.; Cozzarelli, N. R.; Bustamante, C. *Nature* **2003**, 424, 338-341.
15. Mirzabekov, A. D.; and Rich, A. *Proc. Natl. Acad. Sci .U S A.* **1979**, 76, 1118-21.

16. Manning, G. S.; Ebralidse, K. K.; Mirzabekov, A. D.; Rich, A. *J Biomol Struct Dyn*. **1989**, 6, 877-89.

17. Strauss, J. K.; Maher, L. J. 3rd *Science*. **1994** , 266, 1829-1834.

18. Hardwidge, P. R.; Lee, D.; Prakash, T. P.; Iglesias, B.; Den, R. B.; Switzer, C.; Maher, L. J. 3rd *Chem Biol*. **2001**, 8, 967-980, and reference cited therein.

19. Williams, L. D.; Maher, L. J. 3rd. *Annu Rev Biophys Biomol Struct.* **2000**, 29, 497-521.

20. Ross, E. D.; Den, R. B.; Hardwidge, P. R.; Maher, L. J. 3rd. *Nucleic Acids Res*. **1999**, 27, 4135-42.

21. Strauss-Soukup, J. K.; Rodrigues, P. D.; Maher, L. J. 3rd. *Biophys Chem*. **1998**, 72, 297-306.

22. Strauss-Soukup J. K.; Vaghefi, M. M.;, Hogrefe, R. I.; Maher, L. J. 3rd. *Biochemistry.* **1997**, 36, 8692-8.

23. Koo, H. S.; Crothers D. M. *Proc. Nat. Acad. Sci., U.S.A.* **1988**, 85, 1763-1767.

24. Wu, H. M.; Crothers, D. M. *Nature (London)* **1984**, 308, 509-513.

25. Hardwidge, P. R.; Wu, J.; Parkhurst, K. M.; Parkhurst, L. J..; Maher, L. J. *Biochemistry* **2002**, 41, 7732-7742.

26. Hardwidge, P. R.; Lee, D. J.; Prakash, T.; Iglesias, B.; Den, R.; Switzer, C.; Maher, L. J. *Chemistry & Biology* **2001**, 10, 967-980.

27. Hardwidge, P. R.; Pang, Y, P.; Zimmerman, J. M.; Vaghefi, M.; Hogrefe, R.; Maher, L. J. in Mohanty, U. and Stellwagen, N (Eds), *ACS Symposium Series* **2004**, 884, 111-132.

28. Podesta, A.; Indrieri, M.; Brogiolo, D.; Manning, G. S.; Milani, P.; Guerra, R.; Finzi, L.; Dunlap, D. *Biophys. J*. **2005**, 89, 2558-63.
29. Clouter, T. E.; Widom, J. Mol. Cell. **2004**, 14, 355-362.
30. Du. Q.; Smith, C.; Shiffeldrim, N.; Vologodskaia, M.; Vologodskii, A. *Proc. Natl. Acad. Sci. U S A*. **2005**, 102, 5397-402.
31. Ross, E. D.; Hardwidge, P. R; Maher, L. *J. Mol. Cell. Biol*. **2001**, 21, 6798-605.
32. Becker, N. A.; Kahn, J. D.; Maher, J. L. *J. Mol. Biol*. **2005**, 49, 716-30.
33. Ross, E. D.; Keating, A. M.; Maher, L. J. *J. Mol. Biol*. **2000**, 297, 321-34.
34. (a) Odijk, T. *J. Poly. Sci. Poly. Phys. Ed*. **1977**, 15, 477-483. (b) Sivokob, A.; Khrapunov, S. N. *Biophys. Chem*. **1997**, 67, 85-96; see references cited therein.
35. Manning, G. S. *Macromolecules* **2001**, 34, 4650-4655.
36. Wiggins, P. A.; Phillips, R.; Nelson, R. C. *Phys. Rev. E.* **2005**, 71, 021909-19.
37. Manning, G. S. *Biophysical J*. **2006**, 91, 1-10.
38. Manning, G. S. *Biophys. Chem*. **2002**, 101-102, 461-473.
39. Manning, G. S. *Q. Rev. Biophys*. **1978**, 11, 179-246.
40. Landau, L. D.; Lifshitz, E. M. *Statistical Physics*, 3rd Ed., Pergamon Press, Oxford, UK, **1988**.
41. Range, K.; Mayaan, E.; Maher, L. J. III.; York, D. M. *Nucl. Acid. Res*. **2004**, 33, 1257-1268.
42. Cairney, K. L.; Harrington, R. E. *Biopolymers* **1982**, 21, 923-934.

43. Rizzo, V.; Schellman, J. A. *Biopolymers* **1981**, 20, 2143-2163.

44. Borochov, N.; Eisenberg, H.; Kam, Z. *Biopolymers* **1981**, 20, 231-235.

45. Sobel, E. S.; Harpst, J. A. *Biopolymers* **1991**, 31, 1559-1564.

Chapter Two

The Role of Large Thermal Fluctuations and Magnesium Ions in t-RNA Selectivity of the Ribosome

2.1 Introduction

In 1944, Erwin Schrödinger made a prescient conjecture that biological machinery may be constructed in a manner "different from anything we've yet tested in the physical laboratory".[1] Human constructed machines are designed to minimize friction, whereas the molecular machinery of life is embedded in a viscous fluid where dissipation and the associated fluctuations are enormous.[2]

Single molecule methods have allowed us to directly observe fluctuating transitions between intermediate states that help elucidate the interplay between structure, dynamics and function. Through these observations, evidence is growing that bio-molecular machinery may be co-opting thermal fluctuations as a fundamental part of their operation. In our study of the minimal functional structure of the hairpin ribozyme, we suggested that the highly directional, enzymatic cleavage of a sequence specific strand of RNA was driven by reversible transitions between intermediate states driven by thermal fluctuations.[3] Subsequent measurements of the transition rates between intermediate states in the full four-helix junction of the hairpin ribozyme reinforced this conjecture.[4]

In single molecule studies of the ribosome,[5] we proposed that thermal fluctuations between cognate and near cognate t-RNA played a critical role in the selection fidelity of the ribosome.[6] The single molecule studies indicate that base pairing of the codon-anticodon complex involve not only interactions with specific residues and ribosomal proteins, but also conformation fluctuations of the ribosome. These conformation changes bring about the small subunit to enfold

around the ternary complex.[7] As a result, the cognate ternary complex, EF-Tu(GTP)aa-tRNA, swivels into an orientation that allows it to make stabilizing contacts with various functional groups such as GAC and sarcin-ricin loop (SRL) near the GTPase activated state.[5,7]

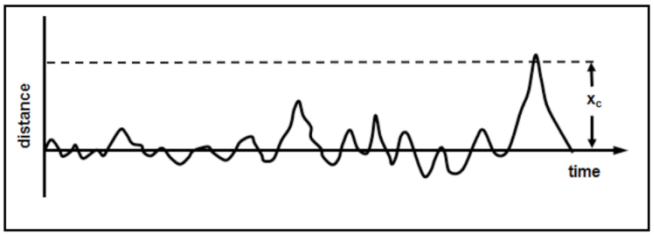


Figure 1. The ternary complex undergoes large thermal fluctuation (denoted by x_c). These fluctuations involve configurational searches that are in the tail end of the probability distribution.

We infer that large and rare thermal fluctuations (Figure 1) of the ribosome are essential to rotate the ternary complex into a position so as to facilitate stable contacts with GAC and the sarcin-ricin loop in the large subunit.[4,5] This inference is based on the time-scale for forming the stable contacts is between 65 ms and 100 ms at 15 mM and 5 mM Mg^{2+} concentrations respectively and the distance traversed by the ternary complex going to the GTPase activated state, the mid-FRET state, is over 70 Å. In this work, a framework to describe the probability of such rare events will be formulated.

We also examine in detail various structural motifs of magnesium binding sites to further elucidate the interplay between structure, dynamics and function

within the ribosome. Divalent metal ions serve an essential role in processes such as DNA bending[8,9] and folding of RNA.[10,11] In particular, magnesium ions play an important role in the subunits association, tRNA binding to the decoding site and in general the structure and the stability of the ribosome.[12,13] As shown in bacterial 70S ribosome, the divalent metal ions interact to hold the ribosomal subunits together.[14] A comparative analysis of the crystal structures of the large subunits in *T. thermophilus*, *E coli,* and *H. marismortuni* ribosomes indicates that around 40% of the magnesium binding sites are conserved not only across the domains of life, but also after the subunits associate.[15,16] In the folded conformation, various phosphate oxygen atoms in the ribosome are less than accessible to the solvent. These phosphate oxygens have an inclination for either inner-sphere or outer-sphere contacts with magnesium. In tertiary folding of ribosomal subunits, magnesium ions are thought to bury the phosphate oxygen atoms that are exposed in unfolded conformation.[16]

The protein functional groups, the structural positioning of the RNA, and the inner-sphere coordination of the protein atoms to them, create several unique motifs for the binding of Mg^{2+} ions in the large subunit of *H. marismortuni* ribosome as well as in *E. coli*.[15,16] Klein *et al*. have identified and structurally classified 116 magnesium binding sites from a high-resolution crystal structure of *H. marismortuni* .[16] The structural classification are labeled by the number of inner sphere contacts (SI text) they form and on the three dimensional arrangement of the protein atoms or the RNA.[16]

Here we present a theoretical framework to determine the probability that as a result of large and rare thermal fluctuations (Figure 1), the cognate ternary complex forms stabilization contacts with the GTPase activated state. We also investigate the binding energies of several unique site-binding Mg^{2+} structural motifs that are a result of the protein functional groups, the structural positioning of the RNA, and the inner-sphere coordination of the protein atoms to them.

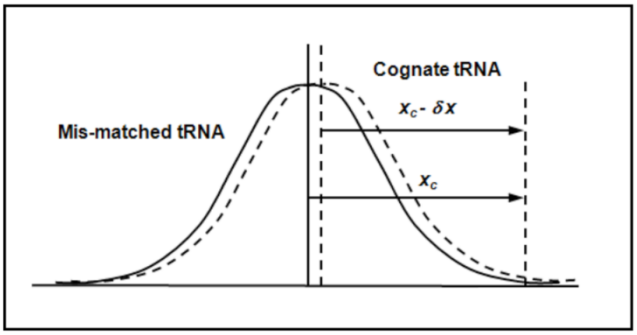


Figure 2. The figure is a plot of the probability density (y-axis) versus coordinate of the ternary complex (x-axis). x denotes the departure of the ternary complex from its equilibrium position. As thermal fluctuations reach a critical value, a near-cognate ternary complex will reach distance x_c that allows it to be in spatial contacts with GAC/SRC sites in the ribosome. Small variations in the re-positioning of cognate relative to near-cognate (mismatched) complexes lead to significant rate enhancement.

2.2 Results

Large thermal fluctuations play a crucial role in the selection process. The probability of forming contacts to stabilize the GTPase activated state involves

configurational searches that are in the tail end of probability distribution (Figure 2). To determine the probability of such an event, let us assume that the thermodynamic state of the 30S subunit and the codon-anticodon complex is described by a set of extensive macroscopic variables $\{\psi_I; I = 1,...,n\}$. The state variables are assumed to be independent, and can be expressed as sums of molecular variables.

We assume, following Onsager and Machlup,[17] that the macroscopic variables defining a state are Gaussian random variables.[18] The probability $P\{\psi_I\}$ of a given fluctuation is proportional to $\exp(S(\psi_1,....\psi_n)/k_B)$, where $S(\psi_1,...,\psi_n)$ is the entropy, which is a function of the macroscopic variables $\{\psi\}$. On expanding the entropy about the equilibrium value, $P(\{\psi\}) \propto e^{-\Sigma ij \psi i (\chi^{-1})_{ij} \psi j}$, where the sums over i and j run over the phase variables, and χ^{-1} is a covariance matrix.

Now consider a fluctuation in component $i = 0$ for which $\Delta\psi_o \geq a$, where a is a positive quantity, and the parameter Δ is small. Since Δ is small, the fluctuations defined by ψ_o are large and rare events. The probability for the rare event occurs in the tail end of the distribution and is given by

$$P(\Delta\psi_o \geq a\} \propto \int_a^\infty dx e^{-\psi^2/2(\Delta^2 \chi_{oo})}. \qquad (1)$$

For simplicity, we have considered a one-dimension probability distribution.

The dominant contribution to the probability is expected to come from those configurations for which $\Delta\psi - a$ is small. By carrying out the integral for $\Delta \square$ 1, one obtains

$$P\{\Delta\psi_o \geq a\} \approx \left[\sqrt{\frac{\chi_{oo}\Delta^2}{2\pi a^2}}\right] e^{-a^2/2(\Delta^2\chi_{oo})}. \qquad (2)$$

The quantity in the square brackets in the above equation dictates the *size* of the fluctuation. The exponential term governs the *decay* of probability of the rare event.

Let $I\{\Phi\}/\Delta^2 = \square/2\Delta^2$, where $\Phi_i = \Delta\psi_i$ and $\square = \Sigma_{ij}\Phi_i\chi^{-1}_{ij}\Phi_j$. The rare event is defined by $\Phi_o \geq a$. Now minimize $\square$ subject to the constraint $\Phi_o \geq a$. Analysis shows that the minimum of I evaluated at the extremum Φ^* is

$$I\{\Phi^*\}/\Delta^2 = a^2/2\Delta^2\chi_{oo}. \qquad (3)$$

Observe that $I\{\Phi^*\}/\Delta^2$ is equal to negative of the exponent that appears in the second term of Eq. (2).

Even though large thermal fluctuations take place in the movement of the ternary complex, our results indicate that the probability of such a rare process is localized around the most likely configuration. This feature is a unique attribute of Varadhan's formulation of stochastic processes of rare events, and in which $I\{\Phi^*\}$ is the action functional.[19-20] The results can be generalized to Gaussian Markov and non-Markovian processes.

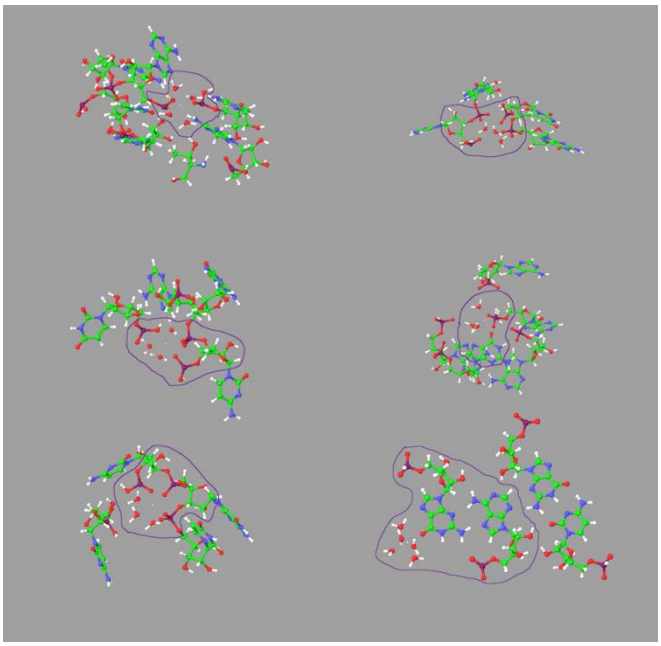


Figure 3. Magnesium binding motifs extracted from high-resolution crystal structure of the 23S RNA subunit of *Haloarcula marismortuni* and from the 70S bacterial ribosome *Escherichia coli*. Mg 10 (top row, left), Mg 7 (top row, right), Mg 8 (second row left), Mg 4 (second row, right), and Mg 26 (third row, left) are from the large ribosomal subunit of *H. marismortui*. A magnesium binding motif along with its interactions in the inter-subunit bridge B3 from the large subunit of *E. coli* 70S ribosome is depicted in the third row (right). The region encircled denotes the quantum mechanical region. The remaining atoms in the motif define the molecular mechanics region. The color code is such that white denotes the hydrogen atoms, red denotes the oxygen atoms, blue denotes the nitrogen atoms, green denotes the carbon atoms, and purple denotes the phosphate group.

Site-bound magnesium ions contribute to structural stability of the ribosome. Which of the magnesium ions observed in the crystal structure of the large subunit of *H. marismortuni* and *E. coli*. ribosome are essential for its structural stability are far from well understood. To provide insights into how site-bound metal ions contribute to structural stability, we have investigated the binding energies of several unique Mg^{2+} binding sites in the ribosome. One such motif (geometrical class IIIa; SI text) is shown in Figure 3 (top row, left). This Mg^{2+} ion (labeled Mg 10) links via inner-sphere the phosphate oxygen atoms in helix 35 and helix 90, to an interface between protein L3 and domains V and II of 23S RNA 23S RNA subunit of *H. marismortuni* ribosome.[16] Mg 10 stabilizes the quaternary as well as the tertiary structures of the 23S subunit of the ribosome.[16]

Magnesium ions are known to bind to small structural motifs in 23S RNA. These ions stabilize the RNA structure by neutralization of part of charge on the phosphate groups. One such case we studied (geometrical class IIa; SI Text) consists of a Mg^{2+} ion, (labeled Mg 27) from the 23S RNA subunit of *H. marismortuni* that facilitates interactions (Figure 4; second row, left) between the phosphate group and the 2-amino of G nucleotide with pro-S_P phosphate oxygen of nucleotide 5' and the 6-oxo group of guanosine base.[16] Here, the magnesium ion forms two inner-sphere coordination with protein residues and RNA bases.

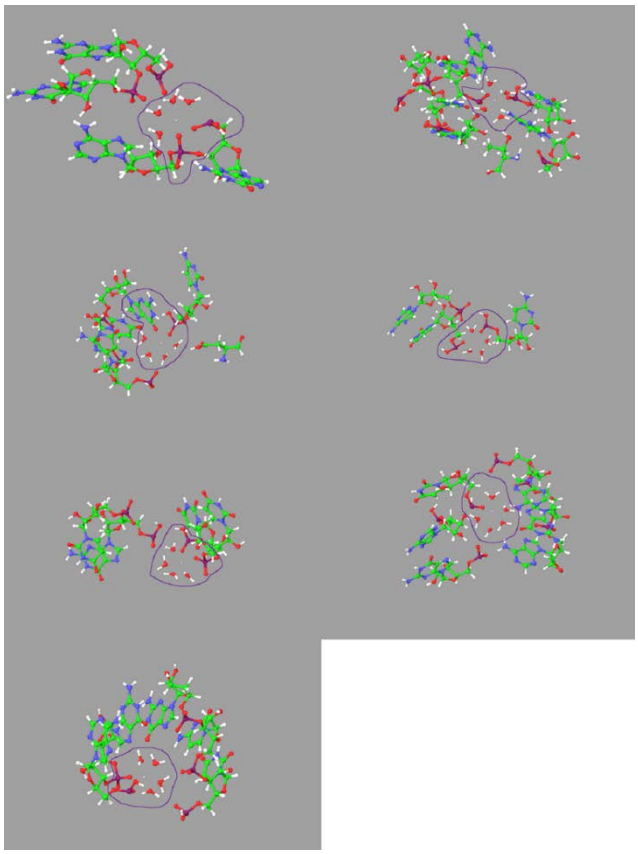


Figure 4. Magnesium binding motifs extracted from high-resolution crystal structure of the 23S RNA subunit of *Haloarcula marismortuni* (PDB:1s72). Mg 9 (top row, left), Mg 15 (top row, right), Mg 27 (second row, left), Mg 29 (second row, right), Mg 30 (third row, left), Mg 39 (third row, right), and Mg 18 (fourth row, left). The region encircled denotes the QM region. The remaining atoms in the motif denote the MM region. The color code is such that white denotes the hydrogen atoms, red denotes the oxygen atoms, blue denotes the nitrogen atoms, green denotes the carbon atoms, and purple denotes the phosphate groups.

Another unique Mg^{2+} ion binding motif occurs in large subunit of *E. coli* 70S ribosome. Most of the observed magnesium binding sites in 30S within 70S are identical to those in the isolated 30S subunit. An exception is a binding site near bridge B3; the later has been proposed to serve as a pivot point for ratchet-like motion of the small subunit.[15] In the 30S subunit, a magnesium ion that adjoins (Figure 3; third row, right) tandem G-A base pairs in helix 44 of 16S rRNA interacts by hydrogen bonding with G-C base pairs in helix 71 of 23S rRNA.[15] Upon association of the 30S and the 50S subunits to form the 70S ribosome, crystal structures reveal that this magnesium ion, instead, forms inner-sphere coordination with nucleotide G1417.[15]

The molecular fragments that define the magnesium binding motifs are extracted from high-resolution crystal structure of the 23S RNA subunit of *Haloarcula marismortuni* (PDB:1s72)[16] and from the crystal structure of h44 of 16S rRNA (PDB:2QB9) and fh71 of 23S rRNA (PDB:2QBA) for the 70S bacterial ribosome *Escherichia coli*.[15] Each complex is partitioned (SI Text) into an inner region and an outer region (Figures 3 and 4).[21,22] The inner region consisting of between 24-56 atoms is treated by *ab initio* electronic structure calculations (QM). All calculations in the QM region are single point and were performed at the DFT-M06-2X/6-31++G** level of approximation (SI Text).[21] The remainder of the atoms (outer region) in the magnesium binding motifs is described by molecular mechanics (MM) and based on the OPLS-2005 force field.[21]

The binding energy, which is the sum of the interaction and the solvation energies, of Mg 10 (type IIIb) is -88.745 kcal/mol while that of Mg 4 (type IIa) to be -73.367 kcal/mol. In contrast, the binding energy of Mg^{2+} ion from *E. coli* 70S ribosome is -47.168 kcal/mol. Other type IIa magnesium ions in the 23S RNA subunit of *H. marismortuni* that we have studied (Figures 3 and 4; SI Table 1) have binding energies -72.983 kcal/mol (Mg 9), -76.625 kcal/mol (Mg 18) and -81.452 kcal/mol (Mg 30). Binding energies of Mg 7 (type IIIb), Mg 26 (type IIIa), and Mg 30 (type I) are comparable. The former two magnesium ions interact with 3 water molecules, while the later with 5 water molecules (SI Table 1).

2.3 Discussion

The ternary complex, as a result of rare and large thermal fluctuations (Figure 1), forms contacts with SRC/GAC leading to stabilization of the GTPase activated state. Our theoretical findings establish that configurations, which make significant contribution to the probability of the rare process, have the inherent property that they are localized tightly around the most likely configuration. We show that this configuration is obtained by minimizing an action functional with an appropriate constraint. The minimized action gives the probability of the rare event.

Small variations in the re-positioning of cognate relative to near-cognate complexes lead to significant rate enhancement (Figure 2). Let x be the departure of the ternary complex from its equilibrium position. As thermal

fluctuations reach a critical value, a near-cognate ternary complex will reach distance a, that allows it to make spatial contacts with GAC/SRC at elongation factor-Tu. In contrast, the critical fluctuation will allow a cognate ternary complex to reach, instead, a distance $a-|\gamma|$, where γ is a constant. From Eq. (2), the ratio of the rate of docking of cognate to the near-cognate ternary complex is approximately

$$(\nu_c / \nu_{nc})(1-|\gamma|/a)^{-1}\exp(a|\gamma|(1-|\gamma|/a)/\chi_{oo;c}\Delta^2), \quad (4)$$

Where the subscripts c and nc stand for cognate and near-cognate complexes, ν is the docking frequency, and we have assumed that $\chi_{oo;c} \approx \chi_{oo;nc}$. Observe that the model predicts that rate enhancement is dictated in part by $\chi_{oo;c}$, the second derivate of the entropy with respect to macroscopic variables.

The large fluctuation theory proposed here is reminiscent of a multidimensional threshold model in chemical kinetics by Slater[23] and in the quantum theory that describes barrier crossing in condensed phases by Wolynes.[24] There are also processes in biological systems where fluctuations in the environment are slower than events that lead to barrier crossings.[25] In these systems, classical kinetic laws are inapplicable since the time dependence of correlation functions display nonexponential behavior. Wang and Wolynes have studied such a system by considering a reaction that occurs when a threshold fluctuation of an environmental variable takes place.[25,26] These authors have

shown that rare events make a significant contribution to the statistical time path of an intermittent process.[25,26]

Whitford *et al.* have recently shown that large thermal fluctuations play a vital function in the process of accommodation in the ribosome.[27] This work supports the mechanism proposed in this paper, namely that small structural changes in the cognate tRNA probablity distribution can make the rare events for the cognate ternary complex more likely to take place than for the non-cognate ones. There is an amplification effect since the events occur at the tail of the distribution.

The nature of interaction of magnesium ions with RNA is generally classified either via site binding or diffuse-binding.[28,29] According to the counterion condensation model, the magnesium ions form a delocalized cloud that is spread over the phosphate backbone.[28,29] These non-specific interactions, i.e., diffuse binding by the magnesium ions lead to substantial neutralization of the negative phosphate charges and thermodynamic stabilization of the RNA.[29] In contrast, in site binding, magnesium ions with its partial layer of water molecules can directly coordinate with RNA.[10,30] Which of the site-bound magnesium ions observed in the crystal structure of the large subunit of *H. marismortuni* and *E. coli*. are essential for its structural stability are not well understood.

The binding energy of various magnesium structural binding motifs in the large subunit of *Haloarcula marismortuni* ribosome is listed in Table 1 (SI Text). In all cases, the total solvation energy is destabilizing while the interaction energy

is stabilizing. A comparison of the gas phase energy and the solvation phase energy of the ligand, i.e., the magnesium ion with its associated water molecules, indicates that many energy values were consistent and predictable for each classified type of binding site (SI Tables 2-3). For type IIa motifs investigated, the gas phase energy of the ligand is around -505.15 Hartree, for example. There were other binding motifs, such as Mg 7 (type IIIb) and Mg 8 (type IIIb), whose ligand gas phase energy as well as solvation phase energy (SI Tables 2 and 3) displayed characteristics of other types, namely Mg 26 (type IIIa) and Mg 16 (type IIIa). Remarkably, the solvation energy also does not vary much within a given type. These trends are also observed at the LMP2 level of approximation for the QM region.

In summary, we have developed a theoretical technique that allows us to determine the probability that the ternary complex forms contacts leading to stabilization of the GTPase activated state. The configurational search for such process is in the tail end of the probability distribution, and is a rare event. The probability distribution of the rare event is localized around the most likely configuration. The binding energies of over a dozen site-bound magnesium structural motifs were investigated and provided insights into the nature of interaction of divalent metal ions with the ribosome and its structural stability.

2.4 Materials and Methods

The magnesium-binding motifs were extracted from the crystal structures[15,16] of *H. marismortuni* (PDB:1s72) and *E. coli* (PDB:2QB9; PDB:2QBA) and capped with hydrogen atoms. The optimization of all hydrogen atoms is implemented by freezing the coordinates of all heavy atoms. All minimizations were carried out in Maestro[21] using a conjugate gradient method in which the energy change and the gradient criterion are taken to be 1.0×10^{-7} kcal/mol and 0.01 kcal/ mol Å, respectively.

Since the number of atoms in the motifs is large, we partitioned the complex into inner and outer regions[22] (see SI Text) using Qsite suite of programs.[21] Quantum mechanical approximations at the DFT-MO6-2X/6-31++G** level and OPLS-2005 force field[21] were employed in the inner and the outer regions, respectively, to obtain the interaction energies in the gas phase (SI Text). The number of atoms in the inner region varied between 24 and 56 depending on the total number of atoms in the magnesium ion structural motif. The interaction energy is, by definition, the difference between single-point energies of the molecular fragments that define the binding motif and the monomers.

We have determined the effects of solvation on the interaction energy by solving the non-linear Poisson-Boltzmann equation using finite element method.[31] The displacement threshold and non-bonded cutoffs were set to 0.1 Å and 12.0 Å. This former signifies how far a given atom moves from the coordinates used in the previous step of the finite element grid calculation of the

non-linear Poisson-Boltzmann equation. Convergence criterion is based on energy change (1.0×10^{-7} kcal/mol) and gradient (0.01 kcal/mol Å). The non-bonded interactions (electrostatic and van der Waals) cutoff is set to 30 Å while the dielectric constant ε in the gas phase is set to unity. Since water is the solvent, the dielectric constant was taken to be a constant, namely 80.4. The temperature and pressure were fixed at 298.15 K and 1 atm respectively. The binding energy is the sum of the interaction and the solvation energies.

References

1. Schrodinger, E. *What Is Life? Cambridge: Cambridge University Press,* **1944**.

2. Chu, S. Is Life Based on the Laws of Physics. in *Visions of Discovery: New Light on Physics, Cosmology, and Consciousness*, eds. Chiao, R.Y.; Cohen, M.L.; Leggett, A.J.; Phillips, W.D.; and Harper, Jr. CL. *Cambridge: Cambridge University Press*, **2010**, pg. 452-470.

3. Zhuang, X.; Kim, H.; Pereira, MJB; Babcock, H.P.; Walter, N.; Chu; S. *Science* **2002**, 296,1473-1476.

4. Nahas, M.; Wilson, T.J.; Hohng, S.; *et al. Nature Struct. Molec. Biol.* **2004**, 11, 1107.

5. Blanchard, S.C.; Kim, H.D.; Gonzalez, Jr. R.L.; Puglisi, J.D.; Chu, S. *Proc. Nat. Acad. Sci. USA,* **2004**,101, 12893–12898.

6. (a) Blanchard, S.C.; Gonzalez, Jr. R.L.; Kim, H.D.; Chu, S.; Puglisi, J.D. *Nature Structural Biology,* **2004**, 11, 1008 – 1014. (b) Lee, T.H.; Blanchard, S.C.; Kim, H.D.; Puglisi, J.D.; Chu, S. *Proc. Natl Acad. Sci. USA,* **2007,** 104, 13661–13665. (c) Gonzalez, Jr. R.L.; Chu, S.; Puglisi, J.D. *RNA,* **2007**, 13**,** 2091–2097.

7. (a) Valle, M.; Zavualov, A.; Li, W. *et al*. *Nat. Struct. Mol. Biol*. **2003**, 10, 899-906. (b) Olge, J.M.; Murphy, F.V.; Tarry, M.J.; Ramakrsihnan, V. *Cell,* **2002**, 111, 721-732.

8. Haran, T,; Mohanty, U. *Q. Rev. Biophys*. **2009**, 42, 41-81.

9. Strauss, J.K.; Maher, 3rd L.J. *Science,* **1994**, 266, 1829-1834.

10. (a) Woodson, S.A. *Curr. Opin. Chem. Biol*. **2005**, 9, 104-109. (b) Pyle, A.M. *J. Biol. Inorg. Chem*. **2002**, 7, 679-690.
11. Schultes, E.; Spasic, A.; Mohanty, U.; Bartel, D. *Nature and Structural Biology,* **2005**, 12, 1130-1136.
12. (a) Kimes, B.W.; Morris, D.R. *Biochemistry,* **1973**,12, 442-449. (b) Chao, F.C. *Arc. Biochem. Biophys*. **1957**, 70, 426-431. (c) Goldberg, A. *J. Mol. Biol*. **1966**, 16, 633-673.
13. (a) McCarthy, B.J. *Biochim. Biophys. Acta*. **1962**, 55, 850-888. (b) Zamir, A.; Miskin, R.; Vogel, Z.; and Elson, D. *Meth. Enzym*. **1974**, 30, 406-426.
14. Selmer, M.; Dunham, C.M.; Murphy, F.V. *et al*. *Science,* **2006**, 313, 1935-1942.
15. Schuwirth, B.S.; Borovinskaya, M.A.; Hau, C.W. *et al*. *Science,* **2005**, 310, 827-834.
16. Klein, D.J.; Moore, P.J.; Steitz, T.A. *RNA,* **2004**, 10, 1366-1379.
17. Onsager, L.; Machlup, S. *Phys. Rev*. **1953**, 91, 1505-1512.
18. Oppenheim, I.; Shuler, K.E.; Weiss, G.H. *Stochastic Process in Chemical Physics: The Master Equation*, *MIT Press: Cambridge*, **1977**.
19. Varadhan, S.R.S. *The Annals of Probability,* **2008**, 36, 397-419.
20. (a) Freidlin, M.I.; Wentzell, A.D. *Random Perturbations of Dynamical Systems*. 2nd ed., Springer, New York, **1988**. (b) Buhler, O. Large deviation theory and extreme waves. *Aha Huliko'a Proceedings*. **2007**, pgs.1-9.
21. Maestro (version 5.5), Jaguar (version 7.6), and Qsite (version 5.5),

Schrödinger, LLC, New York, NY, **2007**.

22. Friesner, R.A.; Dunietz, B.D. *Acc. Chem. Res*. **2001**, 34, 351-358.
23. Slater, N.B. *Theory of unimolecular reactions.* Methuen, London, **1959**.
24. Wolynes, P.G. *Phys. Rev. Lett*. **1981**, 47, 968-971.
25. Wang, J.; Wolynes, P.G. *J. Chem. Phys*. **1999**, 110, 4812-4819.
26. Wang, J.; Wolynes, P.G. *J. Phys. Chem*. **1996**, 100, 1129-1136.
27. Whitford, P.C.; Geggier, P.; Altman, R.B.; Blanchard, S.C.; Onuchic, J.N.; Sanbonmatsu, K.Y. *RNA,* **2010**, 16, 1196-1204.
28. Mohanty, U.; Spasic, A. *Adv. Chem. Phys*. **2008**, 139, 139-176.
29. Manning, G.S. *Q. Rev. Biophy*. **1978**, 11, 179-246.
30. Petrov, A.S.; Lamm, G.; Pack, G.R. *Biopolymers,* **2005**, 77, 137-154.
31. Fogolari, F.; Brigo, A.; Molinari, H. *J. Mol. Recog*. **2002**, 15, 377-392.

Appendix

Files in this Data Supplement:

SI Text

SI Table 1

SI Table 2

SI Table 3

SI Text

Binding energy of various magnesium ion structural motifs in the ribosome

The molecular fragments that define the magnesium binding motifs are extracted from high-resolution crystal structures of the 23S RNA subunit of *Haloarcula marismortuni* (PDB:1s72).[1] The labeling of the magnesium binding sites extracted from *Haloarcula marismortuni* is the same as that introduced by Klein *et al*.[1] In this notation, we have investigated the interaction energy of Mg^{2+} 7, 8, 9, 10, 15, 18 and Mg^{2+} 26, 27, 28, 29, 30, and 39, from *Haloarcula marismortuni* (Figures 3, 4). A structural classification system of numeric types are labeled by the number of inner sphere contacts they form.[1]The magnesium binding motifs are of types I, IIa, IIb, IIIa, and IIIb. Types II and III can be further divided into subgroups based on the three dimensional arrangement of the protein atoms or the RNA. Another magnesium binding motif obtained from the 70S bacterial ribosome *Escherichia coli* is also analyzed.[2] This binding site is

constructed from the crystal structure of h44 of 16S rRNA (PDB:2QB9) and fh71 of 23S rRNA (PDB:2QBA).[2]

Hydrogen atoms were added to the magnesium complex when necessary. The energy optimization of the hydrogen atoms positions is implemented by freezing the coordinates of the heavy atoms to that given by the crystal structure. All minimizations were carried out in Maestro,[3] from Schrodinger suite of programs, using a conjugate gradient method in which the energy change and the gradient criterion are taken to be 1.0×10^{-7} kcal/mol and 0.01 kcal/(mol Å), respectively. OPLS-2005 force field was employed,[4] dielectric constant was set to unity, cutoff distance was taken to be 12.0 Å, and the maximum step size and the cycles for iteration during minimization were 1.000 and 10000, respectively. The minimization is completed in within the range of the maximum iteration cycles used.

As the number of atoms in the various magnesium binding motifs ranges from 100 to over 200, a full quantum mechanical analysis of the interaction energy or the Gibbs free energy is not feasible. Each complex is, consequently, partitioned into an inner region and an outer region (Figures 3 and 4) using Qsite (QM/MM) from Schrodinger suite of programs.[5,6] The inner region consisting of between 24-56 atoms is treated by *ab initio* electronic structure calculations (QM). The remainder of the atoms (outer region) in the magnesium binding motifs is described by molecular mechanics (MM) based on the OPLS-2005 force field for the potential energy.[5,6] In partitioning the complex, one inserts a hydrogen cap on

an atom in the inner region provided this atom is singly bonded to a different atom in the outer region. The interface between the two regions is build using the frozen localized molecular orbitals.[5,7] The inner and the outer regions interact via van der Waals and Coulomb forces.[5,7]

All calculations in the QM region are single point and were performed at the DFT-M06-2X/6-31++G** level of approximations using Qsite suite of programs.[5,8] Convergence is based on energy change (5.00×10^{-5} Hartree) and rms density matrix change (5.00×10^{-6}). In the MM region, we have used the OPLS-2005 force field and frozen the heavy atoms. We have minimized the energy using the conjugate gradient method. Convergence criterion is based on energy change (1.0×10^{-7} kcal/mol) and gradient (0.01 kcal/ mol Å). The non-bonded interactions (electrostatic and van der Waals) cutoff is set to 30 Å while the dielectric constant ε in the gas phase is set to unity. The maximum step size and cycles for iteration during minimization were 1.000 and 5000, respectively.

The number of atoms in the inner region varied between 24 and 56. Lesser atoms for the QM region in some cases were selected for either two reasons. First, there were fewer atoms contained within the specified units of direct binding within the inner sphere. For example, an inner sphere phosphate group will have fewer atoms than say if an RNA base is involved in the direct inner sphere coordination. Second, if a phosphate group was located in between two bases - in this situation, there was no way of avoiding the inclusion of a few

additional atoms because this would lead to violations of Hydrogen capping. In other words, we allow a certain number of adjacent atoms to exist between capping cuts.

In our approach, we only include known inner sphere contacts within the QM region. This approach allows us to analyze binding energy differences between different structural classes. While this resulted in some inner sphere number of atoms to be less than others of a certain type, we believe this to be acceptable, since we are looking to see if energy variations also exist within one type between different magnesium binding sites that contain different inner sphere contacts as well.

Table 1. Binding energies (kcal/mol) for various magnesium ions binding sites in the 23S RNA subunit of *Haloarcula marismortuni*.

Magnesium ion number and number of water molecules (in brackets)	Geometric motif type	Binding energy (kcal/mol)
4 (# water molecules: 4)	IIa	-73.367
7 (# water molecules: 3)	IIIb	-112.486
8 (# water molecules: 3)	IIIb	-71.832
9 (# water molecules: 4)	IIa	-72.983
10 (# water molecules: 3)	IIIb	-88.745
15 (# water molecules: 4)	IIb	-16.341
18 (# water molecules: 4)	IIa	-76.625
26 (# water molecules: 3)	IIIa	-107.805
27 (# water molecules: 4)	IIa	-75.097
29 (# water molecules: 4)	IIa	-16.852
30 (# water molecules: 4)	IIa	-81.452
39 (# water molecules: 5)	I	-104.448

Table 2. Gas phase energy of the ligand (Hartrees) for types IIa and IIb magnesium ions in the 23S RNA subunit of *Haloarcula marismortuni*.

Magnesium ion number	Geometric motif type	Gas phase energy of ligand (Hartree)
9	IIa	-505.155
18	IIa	-505.168
27	IIa	-505.157
29	IIa	-505.165
30	IIa	-505.168
15	IIb	-505.184

Table 3. Gas phase energy of the ligand (Hartrees) for types IIIa and IIIb magnesium ions in the 23S RNA subunit of *Haloarcula marismortuni*.

Magnesium ion number	Geometric motif type	Gas phase energy of ligand (Hartree)
26	IIIa	-428.682
16	IIIa	-428.692
7	IIIb	-428.700
8	IIIb	-428.712

References

1. Klein, D.;Moore, P.B.; Steitz, T.A. *RNA*, **2004**, 10, 1366-1379.
2. Schuwirth, B.S., *et al*. *Science*, **2005**, 310, 827-834.
3. Maestro (version 8), Schrödinger, LLC, New York, NY, **2009**.
4. Kaminski, G.A.; Friesner, R.A.; Tirado-Rives, J.; Jorgensen, W.L. *J. Phys. Chem*. B, **2001**, 105, 6474-6487.
5. Qsite (version 5.5), Schrödinger, LLC, New York, NY, **2009**.
6. Murphy, R.B.; Philipp, D. M.; Friesner, R.A. *Chem. Phys. Lett*. **2000**, 321, 113-120.
7. Philipp, D.M.; Friesner, R.A. J*. Comput. Chem*. **1999**, 20, 1468-1494.
8. Jaguar (version 7.6) and Impact (version 5.5), Schrödinger, LLC, New York, NY, **2009**.

Chapter Three

Probing the α-Helical Structural Stability of Stapled p53 Peptides: Molecular Dynamics Simulations and Analysis

3.1 Introduction

A renaissance of peptide drug discovery has emerged over the past several years. In particular, the identification of synthetically 'stapled' α-helical peptides having promising pharmacokinetic, metabolic stability, and cell-penetrating properties has sparked tremendous interest in their development for a plethora of therapeutic targets that have otherwise been deemed 'undruggable' by more conventional small-molecule strategies.[1] To date, numerous examples of stapled peptides for varying therapeutic targets have been described, including p53,[2] BID BH3,[3,4] BAD BH3,[5] NOTCH,[6] and HIV-1 capsid.[7] Both single-turn (i + 4 -> i) and double-turn (i + 7 -> i) stapling chemistries are represented in these studies.

Synthetic p53 peptides incorporating double-turn stapling chemistry have been previously described[2] and have provided insight to the evolution of a prototype series of cell-penetrating and in vitro biologically effective lead compounds. This series of peptides (Table 1) serves as the basis for the computational work presented here. These stapled p53 peptides have incorporated C_α-methyl-amino acids having terminal olefin alkyl side chains of different lengths and chirality, such as (S)-$(CH_2)_3$-$CH{=}CH_2$ and (R)-$(CH_2)_6$-$CH{=}CH_2$, which upon metathesis form an all-carbon macrocycle via an olefin linkage (see compound 8, Scheme 1). The structure-activity relationships of p53-stapled peptides 1-10 involve further amino acid modifications to decrease negative charge (i.e., Asp and Glu replacement by Asn and Gln, respectively) and facilitate cell penetration as well as point mutations to avoid nuclear export

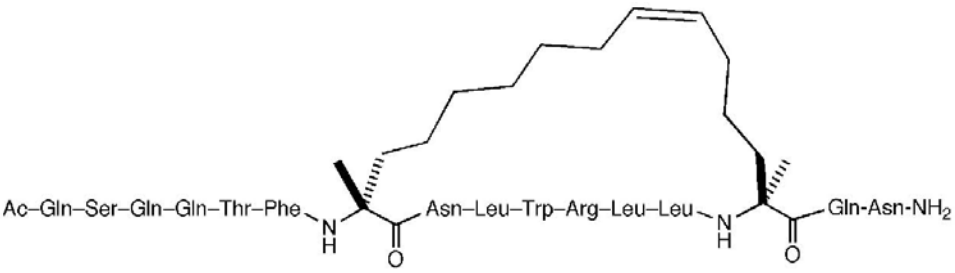
and ubiquitination (i.e., Lys replacement by Arg). In contrast, three key hydrophobic amino acids deemed critical for E3 ubiquitin ligase (MDM2) binding (i.e., Leu, Trp, and Phe as highlighted in stapled p53 peptide analogs 1-10) were conserved throughout the series.

Importantly, the relationship between the structure and dynamic stability of stapled peptides are not well understood. Recently, an all-atom Monte Carlo folding simulation study comparing unmodified peptides derived from RNase A and BID BH3 with varying (I + 4 -> i) and (I + 7 -> i) stapling chemistry has been reported.[8] Another investigation[9] led to the parameterization of minimal, dynamic model reproducing the structural stability profile of a series of BH3-stapled peptides as a function of the cross-linking moiety.

Here, we use extensive molecular dynamics simulations to study a series of stapled α-helical peptides over a range of temperatures in solution. The peptides are found to exhibit substantial variations in predicted α-helicities that are in good agreement with the experimental values.[2] In addition, we find significant variation in local structural flexibility of the stapled peptides with the position of the linker, which appears to be more closely related to the observed differences in activity than the absolute α-helical stability. These simulations provide new insights into the design of α-helical stapled peptides and the development of potent inhibitors of α-helical protein-protein interfaces.

Table 1. The sequences of stapled α-helical p53 peptide analogs

Peptide	Sequence
WT	Ac-LSQETFSDLWKLLPEN-NH_2
SAH-p53-**1**	Ac-LSQETFSDR_8WKLLPES_5- NH_2
SAH-p53-**2**	Ac-LSQER_8FSDLWKS_5LPEN- NH_2
SAH-p53-**3**	Ac-LSQR_8TFSDLWS_5LLPEN- NH_2
SAH-p53-**4**	Ac-LSQETFR_8DLWKLLS_5EN- NH_2
SAH-p53-**5**	Ac-LSQETFR_8NLWKLLS_5QN- NH_2
SAH-p53-**6**	Ac-LSQQTFR_8NLWRLLS_5QN- NH_2
SAH-p53-**7**	Ac-QSQQTFR_8NLWKLLS_5QN- NH_2
SAH-p53-**8**	Ac-QSQQTFR_8NLWRLLS_5QN- NH_2
SAH-p53-**8f**	Ac-QSQQTAR_8NLWRLLS_5QN- NH_2
UAH-p53-**8**	Ac-QSQQTFR_8NLWRKKS_5QN- NH_2



Scheme 1. Structure of stapled p53 analog peptide 8.

3.2 Method and material

(i) Initial Structures

Initial structures of the 16 residue wild type p53 peptide, and the stapled analogs SAH-p53-**1-8** were constructed using Builder module of the Maestro package (Schrödinger) in an α-helical configuration. The alkene bridge was added manually, with the double bond placed in a cis configuration. Hydrogen atoms were added to the peptide as necessary, and the protonation states were assigned to ionizable residues according to the pK_a based on pH = 7.0 using the Protein Preparation workflow in Maestro and the Epik module.[10] The NAC and ACP caps were added to the N and C termini of the peptide, respectively. Each peptide was then placed in a cubic cell, with size adjusted to maintain a minimum distance of 10 Å to the cell boundary, and soaked with a pre-equilibrated box of water using the System Builder module of the Desmond package.[11,12] All overlapping solvent molecules were removed, and an appropriate number of counterions were added to maintain charge neutrality.

(ii) Molecular Dynamics Simulations

All molecular dynamics (MD) simulations were performed using the Desmond package (Schrödinger). The OPLSAA 2005 force field[13,14] was used to model all peptide interactions, and the TIP3P model[15] was used for water. The particle-mesh Ewald method[16] (PME) was used to calculate long-range electrostatic interactions with a grid spacing of 0.08 nm. Van der Waals and short range electrostatic interactions were smoothly truncated at 9 Å. Nose-Hoover thermostats[17] were utilized to maintain the constant simulation temperature and

the Martina-Tobias-Klein method[18] was used to control the pressure. The equations of motion were integrated using the multistep RESPA propagator[19] with an inner time step of 2.0 fs for bonded interactions and an outer time step of 6.0 fs for non-bonded interactions. Periodic boundary conditions were applied throughout. This was followed by 5000 ps of molecular dynamics at constant NPT to equilibrate the system. Following, equilibration the trajectory was then evolved at constant NVT for a further 10 ns, and configurations were saved in 4 ps intervals. The starting conformations harvested from high temperature runs were separately solvated and equilibrated per above protocol prior to the start of low temperature production runs.

(iii) Replica Exchange Molecular Dynamics Simulations

The replica exchange molecular dynamics (REMD) approach[20] involves evolving a number of copies of the system in parallel at different temperatures, and periodically exchanging the configurations between trajectories *i* and *j* with the probability $w_{ij} = \exp(-\Delta)$, where $\Delta \equiv (\beta_i - \beta_j)\ (E_j - E_i)$, and $\beta \equiv 1/k_B T$. This insures proper canonical sampling at all temperatures, with high temperature simulations facilitating barrier crossings, allowing those at low temperature to explore the details of local free energy minima. The systems were prepared and relaxed as described above. In each case 64 replicas of the system were evolved in parallel for 15 ns at constant NVT, with temperatures evenly spaced between 300K and 600K in approximately 4K intervals. After the first 100 ps, replica exchanges between each pair of nearest neighbor trajectories were attempted

every 12 ps. to equilibrate the system. Following a 5 ns period, the configurations were saved in 1 ps intervals. over the final 10 ns of each simulation, providing an aggregate 640 ns of sampling for each system. The exchange acceptance ranged from 22% to 55%, and convergence was confirmed by comparing ensemble averages computed separately over the fist and latter half of the production period.

3.3 Results and Discussion

We have performed extensive all-atom explicit solvent molecular dynamics simulations of a series of 10 stapled p53 peptide analogs (SAH-p53) previously synthesized and studied by Bernal et al.,[2] as well as the wild-type peptide. The stapled peptides contain modifications at two alpha positions in the 16-residue p53 sequence. The nature of the modifications included replacing the side chain with a methyl and a six (S6) or eight (R8) carbon olefin side chains. Stapling is achieved through linking the side chains via a ruthenium catalyzed olefin metathesis. This results in an unsaturated carbon cross-link that spans the modified residues of the peptide. Peptides **1-4** differ from the wild-type peptide in the relative location of the cross-link within the peptide sequence whereas peptides **5-8** have side chain differences, with aspartic acid and glutamic acid being replaced by asparagine and glutamine, respectively. The amino acid replacements were made in an effort to facilitate membrane penetration through the increase in the net positive charge of the peptide, which was achieved

through point mutations of amino acids known to engage in nuclear export and ubiquitylation.[2] In each case, the three hydrophobic amino acids deemed critical for MDM2 binding (Leu26, Trp23 and Phe19) were left unaltered as found in the wild-type peptide sequence. Finally, a version of SAH-p53-8 peptide with the link unclosed and a point mutation F19A completes the series. Table 1 summarizes the names and sequences of the peptides used in this work.

The goal of this study was to determine how and to what extent the introduction of hydrocarbon cross-links at various locations in the sequence modifies the equilibrium conformational population of the peptides in solution. In particular, a crystal structure of the amino-terminal domain of MDM2 protein bound to the transactivation domain of p53, which the stapled peptides seek to mimic, shows the latter to be in a mostly α-helical conformation when bound to MDM2. This structure allows for three hydrophobic amino acids on the hydrophobic side of the helix to adopt extended side-chain conformations and pack with each other. This forms three sequential side-chain contacts with a deep hydrophobic pocket of the MDM2 binding cleft. Circular dichroism (CD) measurements of free SAH-p53 in aqueous solution show a positive correlation between the degree of α-helical content and binding affinity of the peptides for hMDM-2. However, the α-helical content was found to vary widely across the series and several of the peptides with low helical content exhibited significant activity in the low to mid nanomolar range, suggesting a more complex relationship between the peptide structures and binding activity.

Structural distribution

Molecular dynamics simulations offer a way to gain unique and valuable insights into the structure and dynamic behavior of biomolecules at an atomistic level of detail. One of the challenges of molecular dynamics of biomolecules is fully exploring the energetic landscape of the system. Biomolecular have complex potential energy profiles and simulations often get trapped for long periods of time in local minima, thereby failing to sufficiently explore the configurational space. In addition to not observing certain structures that may be biologically relevant, inadequate sampling also prevents recovering the correct thermodynamic distribution that is needed to correctly predict the energetics of the system. In order to improve sampling of the equilibrium ensemble, the following strategy was employed. After an initial setup and equilibration period, each peptide was simulated at a high temperature (500K) where the kinetic energy is sufficiently large to allow easy crossing of the barriers separating the minima. A number of configurations (10-20ns) were selected from each high temperature trajectory and used as a starting point for a moderately long (10 ns) production simulation at room temperature (300K), providing a total of 100-200 ns of sampling for each peptide. The active domain of the p53 is observed to adopt a largely helical structure when bound to the MDM2 receptor. To better understand the structural profile of the peptides in the unbound state we first investigate the helical content of the stapled peptides in solution. Peptides that are not helical in solution would have to incur a conformational penalty to adopt

the helical bound state conformation and therefore are hypothesized to be less active. Furthermore, the degree of helical content in the unbound state should correlate with binding activity if the conformational strain energy in the bound state is a limiting factor in the binding process.

There are numerous ways to define helical structure of the peptide chain. For example, proteins residues are typically classified as helical if their backbone dihedral angles φ and ψ lie in a narrow region of the Ramachandran diagram, with φ between -95° and -35°, and ψ between -15° and -70°. For α-helices, which are the predominant helical form in proteins, the sum of the φ and ψ angles on adjacent residues is ~-105°. This confines the dihedrals to a narrow strip with a slope close to negative one around an ideal α-helix characterized by (φ,ψ)=(-60°,-45°). In contrast, for the two other commonly occurring forms of helices the sums of the above angles are ~-75° for the π-helix and ~-130° for the 3_{10}-helix. Helical conformations are also recognized by a specific pattern of hydrogen bonds formed between the backbone amide (NH) of residue *i* and the carbonyl (C=O) of another residue, where *i*+3 is a 3_{10}-helix, *i*+4 is an α-helix, and *i*+5 is a π-helix. These definitions were used to analyze the conformational ensembles of stapled p53 peptide analogs.

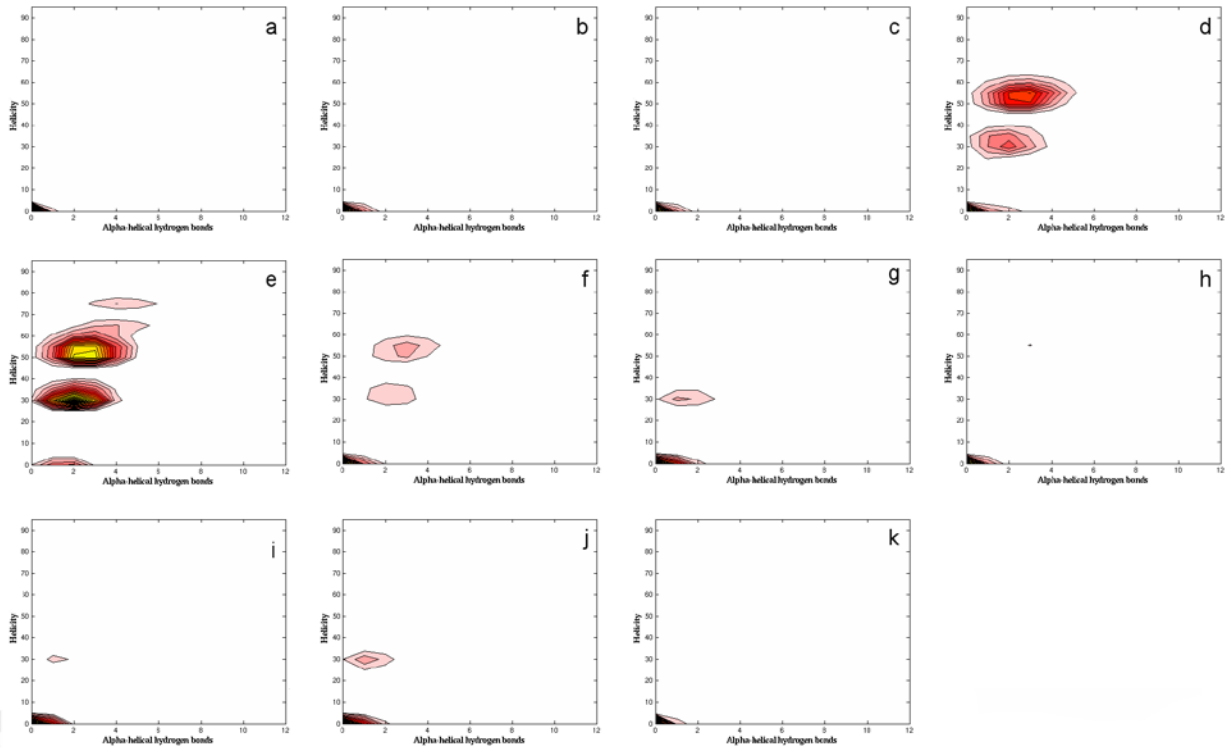


Figure 1. The contour plots of the conformational distribution of the p53 peptide analogs as a function of 39 the number of backbone i→i+4 α-helical hydrogen bonds and the percent α-helical content, for (a) **1**, (b) **2**, (c) **3**, (d) **4**, (e) **5**, (f) **6**, (g) **7**, (h) **8**, (i) **9**, (j) **10**, (k) wild type. Brighter color indicates higher intensity.

In Figure 1 we plot the conformational population distribution as a function of percent α-helical content defined as the ratio of the number of helical residues to the total number of residues in the chain, and the number of α-helical (*i*→*i*+4) hydrogen bonds. The α-helical nature of individual residues was assigned by the STRIDE program.[21] The existence of hydrogen bonds was the determined via a geometry criterion requiring a minimum donor-acceptor distance of 3.5 Å and a maximum A-D-H angle of 30°. Interestingly, the helical population is predicted to

be quite low for most peptides. In particular, SAH-1, SAH-2, SAH-3, SAH-8 and the native p53 show a single pronounced peak at < 5% helicity and 0-1 native hydrogen bonds, indicating minimal α-helical population. SAH-4 and SAH-5 show two additional pronounced peaks, including a broad maximum centered at 30% helicity and 2 hydrogen bonds, and 50% helicity and 3 hydrogen bonds respectively. These peaks are present to a lesser extent in SAH-6, and a small peak at 30% helicity and 1 hydrogen bond is observed for SAH-8f and the unstapled peptide UAH-8.

Table 2. The average helicity and number of α-helix hydrogen bonds

Peptide	Helicity	α-helix H-bonds
WT	2.0%	0.17
SAH-p53-**1**	4.0%	0.44
SAH-p53-**2**	0.0050%	0.17
SAH-p53-**3**	1.8%	0.26
SAH-p53-**4**	38%	2.0
SAH-p53-**5**	45%	2.5
SAH-p53-**6**	26%	1.6
SAH-p53-**7**	15%	1.0
SAH-p53-**8**	22%	1.3
SAH-p53-**8f**	12%	0.85
UAH-p53-**8**	11%	0.72

*The experimental helicities of p53 peptides were obtained through Circular Dichroism spectroscopy data by Verdine *et al.* In which, compounds were dissolved in H_2O, and the concentrations of compounds ranged from 10-50 μM at 20º C.[2]

The average percent helicity for the entire population as calculated by STRIDE is given in Table 2. These are significantly lower than the helical content reported by Verdine et al., which was estimated from CD measurements. One possibility for such a discrepancy may be due to a distortion of the helical structure away from the ideal α-helix geometry, indicating that the helix criteria used in the calculations above may be too stringent for direct comparison to

experimental CD spectra. For example, even the peptides with a significant helical population as predicted by STRIDE still show a relatively low number of characteristic ($i \rightarrow i+4$) hydrogen bonds. To investigate this further, we studied the distribution of the (φ,ψ) dihedrals for the 11 peptides, as shown in Figure 2.

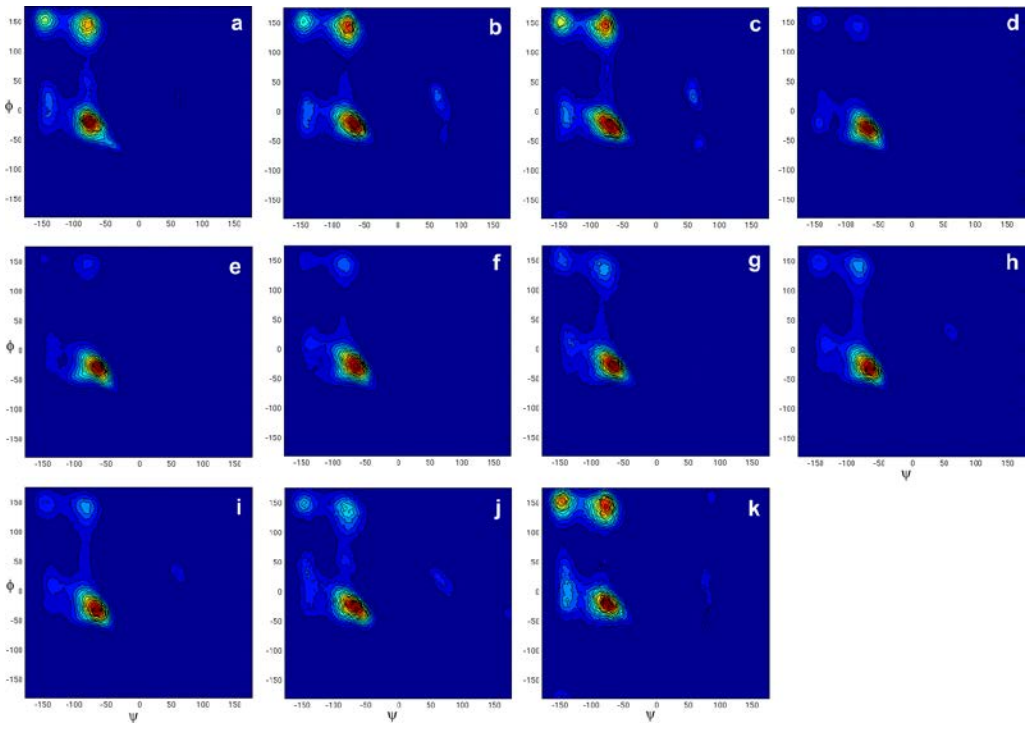


Figure 2. The Ramachandran plot of the distribution of the (φ, ψ) backbone dihedrals for (a) 1, (b) 2, (c) 3, (d) 4, (e) 5, (f) 6, (g) 7, (h) 8, (i) 9, (j) 10, (k) wild type. More red color indicates higher intensity.

All distributions show a strong peak in the range φ=[-95,-50°] and ψ=[-50°,-10°], which is generally the allowed α-helix region of the Ramachandran diagram. Furthermore, these peaks display an oblong shape and lie roughly along a line with a slope of negative one in the (φ,ψ) plane, indicative of a periodic helical

structure. The peptides known experimentally to have weak helical propensity, including SAH-1, SAH-2, SAH-3, and the wild type show significant populations in the two peaks located in extended β-sheet region (the upper left region of the Ramachandran plot), while strongly helical peptides such as SAH-4, and SAH-8 show most of the populations around the α-helix region. Still, the peaks are significantly broader than expected for an α-helix, indicating a distortion of the helical structure from ideal values.

In Table 3 we report the percentage of residues whose dihedral (φ,ψ) pairs were found to lie within a 30° band between (φ,ψ) = (-95°,-15°) and (-35°,-70°), characteristic of an α-helix. The percentages are generally higher than those reported by STRIDE, particularly for the weakly helical peptides and more in line with experimentally measured CD results, supporting the view that the default helical criteria in STRIDE is too stringent for direct comparison with CD spectra and that the helical structures adopted by the peptides are somewhat distorted from their ideal values. We also report the total number of backbone hydrogen bonds, which includes NH – C=O donor acceptor pairings such as those from the 3_{10}-helices, π-helices, and β-hairpin turns, in addition to the α-helical backbone hydrogen bonds. The significantly larger values indicate a considerable contribution from the non-α-helical structural motifs to the equilibrium ensemble.

Table 3. Fraction of backbone dihedrals in the helical region and the average number of backbone hydrogen bonds. Numbers in parentheses are only the α-helix hydrogen bonds (*i* to *i* + 4).

Peptide	α-helical	backbone
WT	14%	2.8 (0.2)
SAH-p53-	20%	2.9 (0.2)
SAH-p53-	17%	2.9 (0.2)
SAH-p53-	17%	3.1 (0.1)
SAH-p53-	40%	4.7 (1.2)
SAH-p53-	44%	5.6 (0.2)
SAH-p53-	35%	4.6 (0.6)
SAH-p53-	28%	4.1 (1.1)
SAH-p53-	32%	4.2 (2.0)
SAH-p53-	24%	3.6 (0.3)
UAH-p53-	24%	3.8 (1.5)

*The experimental helicities of p53 peptides were obtained through Circular Dichroism spectroscopy data by Verdine *et al.* In which, compounds were dissolved in H_2O, and the concentrations of compounds ranged from 10-50 μM at 20º C.[2]

Next, we examine the conformational distribution of the stapled peptides as a function of temperature. This is useful because the degree of persistence of structural features in the peptide across a range temperature provides insights into the stability of secondary structure and the energetic profile of the transition from structured to unstructured states. In Figure 3 we plot the average α-helicity,

as computed by STRIDE, as a function of temperature. The conformations of the peptide were harvested from separate 10 ns constant NPT simulations, following a 5 ns equilibration period performed at each temperature, ranging from 300K to 500K (see Materials and Methods for details).

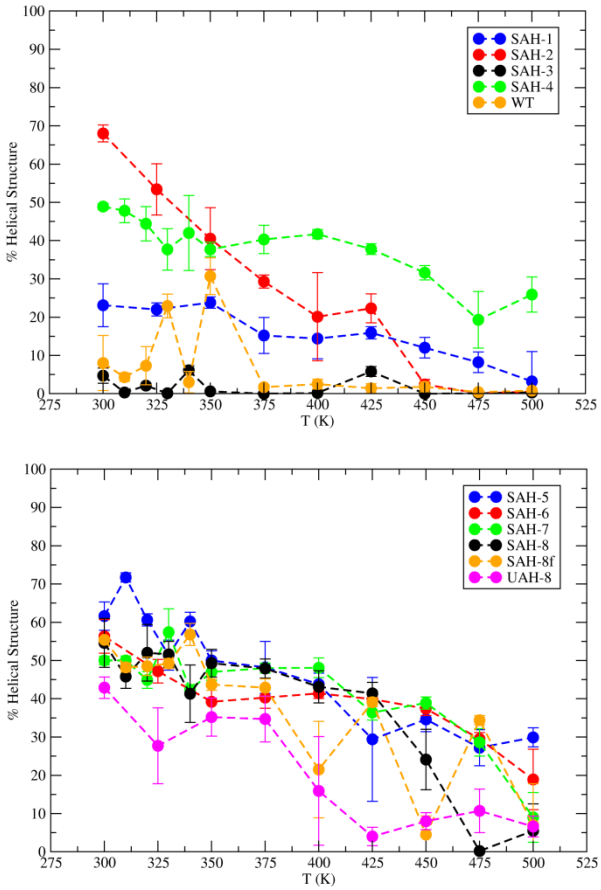


Figure 3. The helix melting curves showing the percent helical content as a function of temperature for (top) **1**, **2**, **3**, **4** and the wild type; (bottom) **5**, **6**, **7**, **8**, **9**, **10**. The color coding is indicated in the legends.

The starting configuration in each simulation was an idealized α-helix. The comparison of the predicted α-helix melting curves shows that peptides with experimentally low helicity, including the wild-type p53 peptide and SAH-1, display virtually complete melting of the helix even at low temperature. SAH-2 is predicted to have a relatively high helicity at room temperature, which is in disagreement with the low helicity observed in the experimental CD spectrum. However, the predicted melting curve decays rapidly with temperature suggesting that the helical structure is quite unstable for this peptide, and longer sampling would likewise yield a lower helicity at room temperature. On the other hand, SAH-1 peptide displays a flat curve, indicating that a low proportion of α-helical structure is maintained across a broad temperature range, suggesting a more stable, yet not necessarily α-helical, structure. Among the first group of peptides, SAH-4 exhibits the highest helicity at room temperature (~50%), in agreement with experiment and maintains this profile up to 450K, indicating a very stable structure. The second series of peptides, SAH-5 through SAH-8, demonstrate similar helical content at 300K, in the range of 40%-60%, with similar melting curves. In contrast with experimental data, SAH-6, which displays significantly weaker CD signal, exhibits a melting curve similar to the rest of the peptides in the series. SAH-8f and UAH-8 display the fastest decay of helical structure, in good agreement with experiment.

To investigate the effect of the alkene linker on the secondary structure of the peptides, Figure 4 shows the per-residue percent α-helicity, as computed by

STRIDE, for the 11 stapled peptides at several temperatures. Several conclusions can be drawn from the plot. First, the per residue helical profiles decay at a rate consistent with the average melting curves shown in Figure 3. However, helical content is not distributed evenly across the amino acid residues in the peptide. In each case, the region of the peptide with the highest helical content corresponds to the portion of the sequence spanned by the linker. Moreover, as the temperature increases this region consistently maintains proportionately more helical content than the rest of the sequence. Hence, it appears that the linker does induce helical structure formation but the effect is primarily local. A similar effect was observed by Kutchukian et al.[8] in their Monte Carlo study of the conformational distribution of stapled peptides derived from RNase A using a knowledge-base pseudo-potential, with the authors attributing the increased helix propensity of the spanned region to the stabilization of the folded state relative to the denatured conformation ensemble. This may explain the significant variations in the potency of the stapled peptides with linkers spanning different portions of the sequence and the lack of direct correlation between the experimental CD spectra and the peptide activity. Interestingly, Kutchukian et al. found that in case of a different sequence of stapled peptides derived from BID BH3, the spanned region is not predominantly helical, and a different, partially folded state is stabilized, indicating that the specific amino sequence of the peptide plays a significant role in determining the preference for helical conformations in conjunction with the location and length of the linker. We

have observed such behavior in the helical propensity of **9**, which is halved with respect that of **8**, even though the only difference is the replacement of a phenylalanine at position 8 with alanine, while the position of the staple and the spanned segment of the sequence is unaltered. Similarly, we observe that in case of peptides with most stable helices such as **8** and **4**, the helical region extends well past the segment spanned by the linker indicating a more comprehensive stabilization. This parallels the findings of Hamacher et al.[9] who, using a simplified Go model of the BH3 peptide, uncovered evidence of global dynamic stabilization of the cross-linked peptide helical structure in addition to the local stabilization imparted by the linker.

Finally, to explore whether the above findings would be maintained with substantially more sampling, a subset of the peptides were subjected to an aggregate 640 ns of replica exchange molecular dynamics (REMD), as described in the Materials and Methods section. The computational resources required to run these simulations limited the number of peptides that could be run via REMD. Figures 5-8 show the helical conformational distribution, Ramachandran plots, per-residue helicity and the helix melting curves obtained from REMD simulations of peptides 4, 8, and the wild type. Comparing Figures 5 and 6 to Figures 1 and 2, it can be seen that there are both similarities and differences. There is consistency in the overall backbone dihedral distribution, which suggests that the conclusions from the shorter, non-REMD simulations have enough sampling to generate reasonable predictions of α-helical content. However, the REMD

simulations for peptide 8 show results that are more in line with the experimental values for α-helical content. Moreover, the conformational distribution of peptide is now also observed to exhibit a multi-peak structure, indicating the presence of two partially folded states in addition to the full α-helical one.

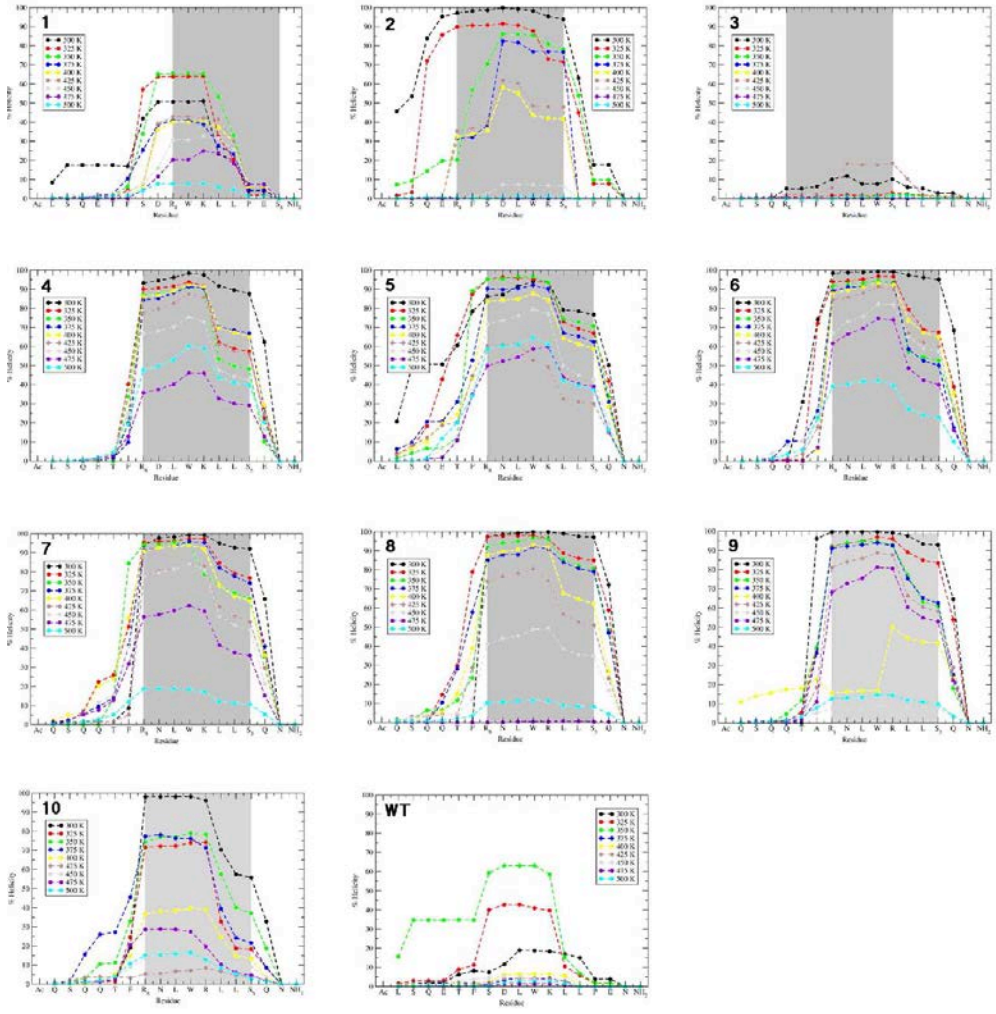


Figure 4. Percent helical occupancy for each of the 16 residues of the p53 peptide analogs, as a function of temperature for peptides **1**-**10** and the wild type p53. The data corresponding to a particular temperature are color coded as given

in the legends. The shaded area indicates the residues spanned by the staple.

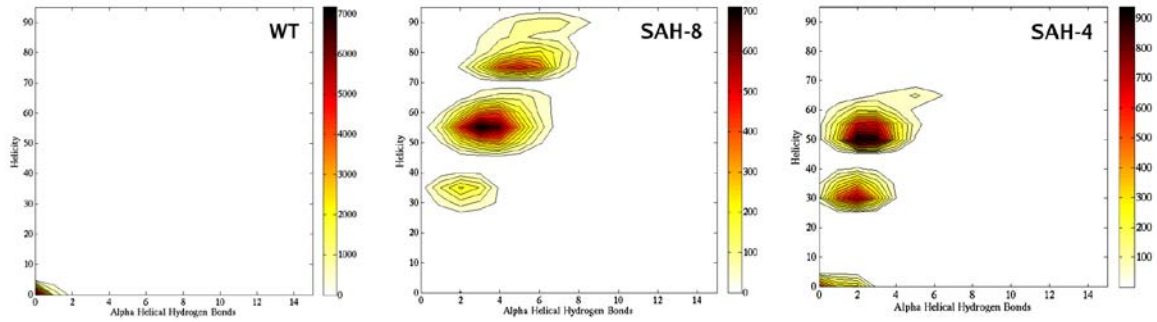


Figure 5. The contour plots of the conformational distribution of the wild type p53 (left), 8 (middle) and 4 (right) peptides, as a function of the number of backbone i→i+4 α-helical hydrogen bonds and the percent α-helical content; computed from the REMD data. The Ramachandran plot of the distirbution of the (φ,ψ) backbone dihedrals.

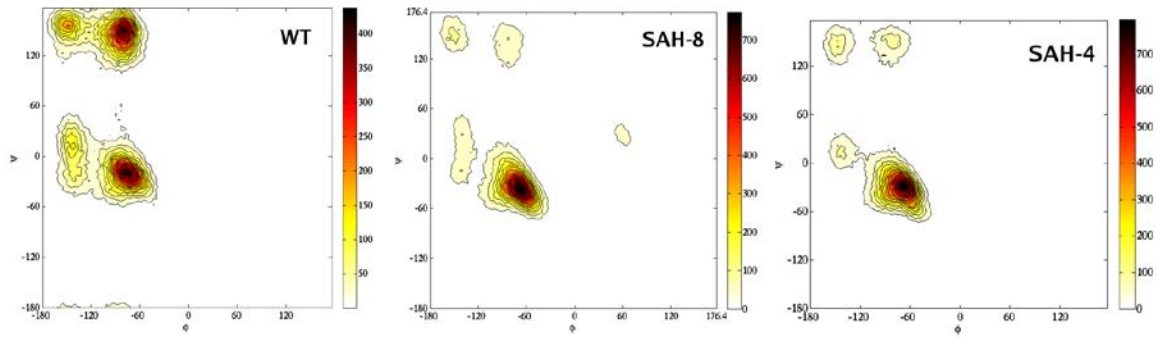


Figure 6. The Ramachandran plot of the distribution of the (φ,ψ) backbone dihedrals for the wild type p53 (left), 8(middle) and 4 (right) peptides, computed from the REMD data.

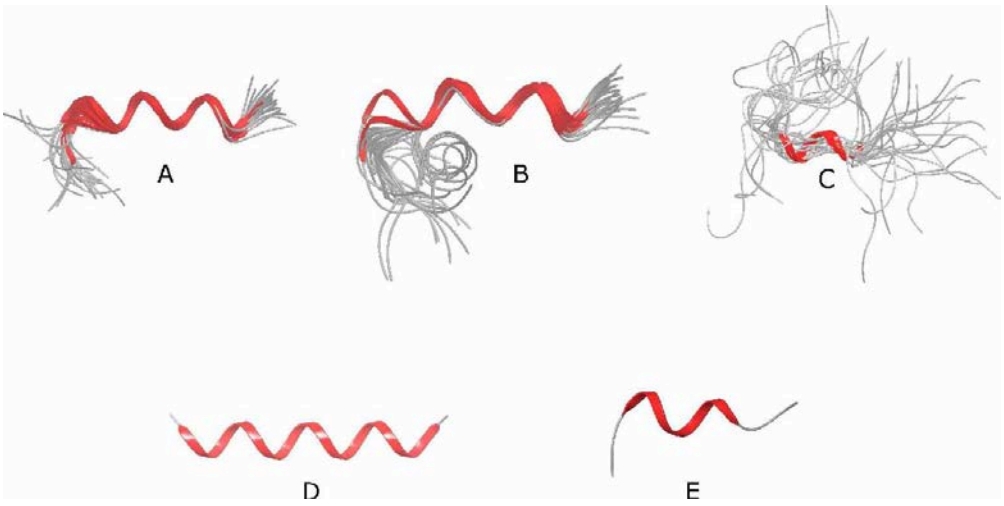


Figure 7. Backbone structures of peptide **8** sampled from the high helicity (A) and medium helicity (B) population cluster, and those of the wild type p53 peptide (C), shown in comparison with the idealized p53 helix (D) and the conformation of p53 in crystal structure of the complex with MDM2 (E).

The melting curves in Figure 8 and the per-residue α-helicity shown in Figure 9 for the REMD simulations are qualitatively similar to the data from shorter simulations, but quantitatively more in line with experimental results. This suggests that for qualitative understanding of α-helical content and relative helical content between peptides it is sufficient to run shorter simulations. However, for quantitative predictions or comparisons with experiments it is necessary to run long simulations, potentially with enhanced sampling, as in the REMD simulations performed in this work.

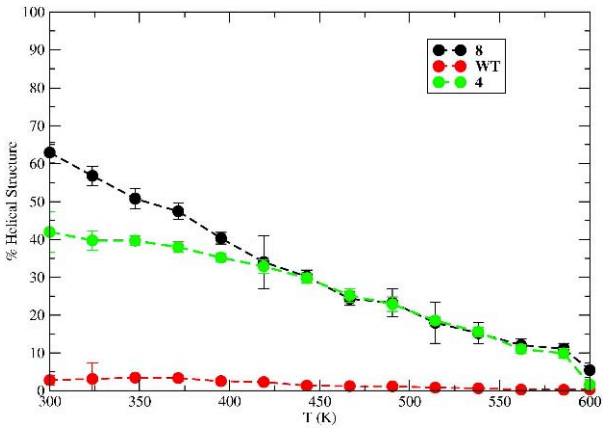


Figure 8. The helix melting curves computed from the REMD data, showing the percent helical content as a function of temperature for the wild-type p53 peptide, **8**, and **4**.

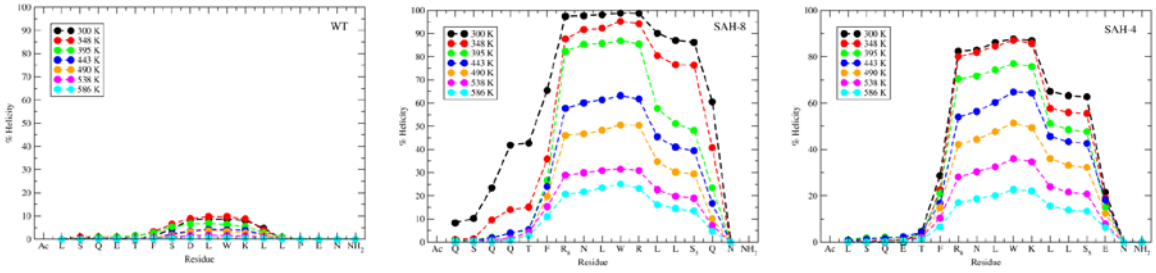


Figure 9. Percent helical occupancy for each of the 16 residues of the wild type (left), 8 (middle), and 4 (right) peptide analogs as a function of temperature, computed from the REMD data. The profiles corresponding to a particular temperature are color coded as shown in the legends.

3.4 Conclusions and future directions

We have used extensive all-atom molecular dynamics simulations to study

the structural distribution of a number of cross-linked analogs of the 16-residue transactivation domain of the p53 protein. The overall propensity to form α-helical structures was found to vary widely across [48 49] the 11 peptide series, and to depend both on the relative position of the cross-linking staple in the peptide sequence as well as the specific amino-acid substitutions. In general, all stapled peptides except for **2** were found to have a higher helical content than the wild-type p53 peptide, which is in agreement with experimental CD measurements. The relative α-helical propensity of most peptides was found to be in good qualitative agreement with experimental measurements, although our values are generally lower than those reported by Bernal et al..[2] This is partly because helical structures formed by the stapled peptides are somewhat distorted from the ideal α-helix geometry, and thus sometimes missed by the helix recognition algorithm in STRIDE. The latter is evident from the Ramachandran plots that show significant, but somewhat broader peaks in the α-helix region of the diagram. Computing the fraction of residues with (φ,ψ) dihedrals within narrowly defined α-helical parameters gives helicities that are closer to the values estimated from CD measurements. REMD simulations performed for several peptides show that improved sampling results in predicted helicities that are very close to experimental values.

Interestingly, simulations of peptides 5 and 6 predict the helical content to be somewhat higher than suggested by experiments. While these peptides share the same position of the crosslink as 4, they differ by substitution of several

acidic residues with their amide counterparts, altering the total charge of the peptide. It is possible that the altered interactions with the specific solvent medium used in CD experiments reduce the helical propensity, which is not fully captured by our simulations. We also found that for several peptides, including 4, 5, 6 and 8, multiple partially-folded states are populated at room temperature. Stabilization of such 'decoy states' by the staples was also observed by Kutchukian et al.,[8] who pointed out that removal of these states may lead to more stable α-helical structures. This is particularly evident in the case of peptide 8, where the most populated state is the partially folded state with mean helicity of 55% (Figure 5). Removal or destabilization of this state would likely lead to a population shift to a nearly fully helical state at mean helicity 75%.

The helical propensities were not distributed equally over the entire sequence. In most cases, the residues with the highest degree of helicity were found to be those spanned by the staple. In a few cases, particularly for 5 and 8, significant helical content was observed in several residues outside of the linked region, while others with the identically positioned cross-link such as 4 do not share this trait. On the other hand, a distinct shoulder is observed in the per-residue helicity distribution for all peptides containing the LL motif in the spanned region. This suggests that, while helix formation is primarily due to local stabilization of helical geometry by the rigid linker, helical propensity is also strongly modulated by the details of the amino acid sequence of the peptide itself, and this should be taken into account in the efforts to design a bioactive molecule. An open question is to

what extent the helical propensity is affected by the nature of the linker -its relative flexibility or rigidity, and conformational preference. We plan to investigate this relationship between the sequence driven and linker induced helical propensity in a future study. Finally, no attempt was made to correlate the properties of the peptides in the solution simulations with binding activity to MDM2 or to simulate the peptides in the bound state. This is also part of future work that will aid in the understanding and design of stapled peptides that enhance binding activity.

References

1. Walensky, L.D.; Verdine, G.L. *Clin Cancer Res*, **2007**, 13, 7264-7270.
2. Bernal, F.; Tyler, A.F.; Korsmeyer, S.J.; Walensky, L.D.; Verdine, G.L. *J. Am. Chem. Soc*. **2007**, 129, 2456-2457.
3. Walensky, L.D.; Kung, A.L.; Escher, I.; Malia, T.J.; Barbuto, S.; Wright, R.D.; Wagner, G.; Verdine, G.L.; Korsmeyer, S.J. *Science*, **2004**, 305:1466-1470.
4. Walensky, L.D.; Pitter, K.; Morash, J.; Oh, K.J.; Barbuto, S.; Fisher, J.; Smith, E.; Verdine, G.L.; Korsmeyer, S.J. *Mol Cell*, **2006**, 24, 199-210.
5. Danial, N.N.; Walensky, L.D.; Zhang, C.Y.; Choi, C.S.; Fisher, J.; Molina, A.J.A.; Datta, S.R. et al. *Nat. Med.* **2008**, 14, 144-153.
6. Moellering, R.E.; Cornejo, M.; Davis, T.N.; Del. Bianco, C.; Aster. J.C.; Blacklow, S.C.; Kung, A.L.; Gilliland, D.G.; Verdine, G.L.; Bradner, J.E. *Nature*, **2009**, 462, 182-188.
7. Bhattacharya, S.; Zhang, H.; Debnath, A.K.; Cowburn, D. *J. Biol. Chem.* **2008**, 283:16274-16278.
8. Kutchukian, P.S.; Yang, J.S.; Verdine, G.L.; Shakhnovich, E.I. *J. Am. Chem. Soc.* **2009**, 131, 4622-4646.
9. Hamacher, K.; Hübsch, A.; McCammon, J.A. *J. Chem. Phys.* **2006**, 124, 164907(1-8).
10. Shelley, J.C.; Cholleti, A.; Frye, L.L.; Greenwood, J.R.; Timlin, M.R.; Uchimaya, M. *J. Comp. Aided. Mol. Des.* **2007**, 21:681-691.

11. Bowers, K.J.; Chow, E.; Xu, H.; Dror, R.O.; Eastwood, M.P.; Gregersen, B.A.; Klepeis, J.L.; Kolossvary, I.; Moraes, M.A.; Sacerdoti, F.D.; Salmon, J.K.; Shan, Y.; Shaw, D.E. Scalable algorithms for molecular dynamics simulations on commodity clusters. In SC '06. *Proceeding of the 2006 ACM/IEEE conference on Suppercomputing*. New York, USA, **2006**.
12. Desmond v2.2, Schrödinger, Inc,: Portland, OR.
13. Jorgensen, W.L.; Maxwell, D.S.; and TiradoRives, J. *J. Am. Chem. Soc.* **1996**, 118:11225-11236.
14. Kaminski, G.A.; Friesner, R.A.; Tirado-Rives, J. *et al* () *J. Phys. Chem. B,* **2001**, 105, 6474-6487.
15. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D. *et al* () *J. Chem. Phys.* **1983**, 79, 926-935.
16. Essmann, U.; Perera, L.; Berkowitz, M.L. *et al*, *J. Chem. Phys.* **1995**, 103, 8577-8593.
17. Hoover, W.G. *Physical Review A,* **1985**, 31, 1695-1697.
18. Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Computational Phys.* **1994**, 101, 4177–4189.
19. Humphreys, D.D.; Friesner, R.A.; Berne, B.J. () *J. Phys. Chem.* **1994**, 98, 6885-6892.
20. Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, 314:141.
21. Frishman, D.; Argos, P. *Proteins,* **1995**, 23: 566-579.

Chapter Four

Conformation and structural stability of all-hydrocarbon *cis* and *trans* linkers single and double stapled to p53 bound to MDM2

4.1 Introduction

As a consequence to DNA damage, transcription factor p53 commences arrest of the natural cycle of a cell.[1,2] The level of cellular p53 is governed by its interactions with E3 ubiquitin ligase, MDM2.[1,2] Over-expression of MDM2 in a cell results in the loss of p53 activity.[1,2]

Proteins often fold into a distinct conformation. However, parts of a protein are less susceptible to fold when they are separated from the rest of the macromolecule.[3] These fragments are frequently found to be in disordered states.[3] Cross-linking a peptide is a novel strategy to avoid unfolding. Specifically, stabilization of a 16-residue helical domain in p53 was recently accomplished by introducing unnatural amino acids with either S or R stereochemistry across one or two turns of a peptide sequence from RNase A, and cross-linked with all-hydrocarbon tethers of different lengths by ring-closing olefin metathesis.[4-6] Circular dichroism studies indicate substantially improved α-helicity. In particular, the helicity of a stapled p53-peptide, *cis*-sah8, whose sequence is shown in Table 1, was 65%, while that of the wild type was 11%.[4,5] Remarkably, *cis*-sah8 slowed cancer cells *in vivo* by turning on apoptotic paths.[4]

In this chapter, the extent to which the tendency for helical configuration of various stapled peptides (Table 1) in solution is modified by the characteristics of its conformation preference when bound to MDM2 has been investigated. An all-hydrocarbon cross-link with an α-methyl group at each terminus were inserted into α-helical peptide scaffolds at positions i and i+7 that span two turns of the

helix. We have also designed a double-stapled peptide by introducing a second hydrocarbon bridge with S configured amino acids incorporated at positions i and i+4 with eight carbon atoms in the cross-link; this staple is on the opposite side of the helix on which the i and i+7 staple resides (Appendix). We varied the sequences of the wild type p53 peptide (Table 1) so as to decrease the net negative charge from -2 to +1, but the triad of amino acids, Phe19, Trp23 and Leu26, that are known to be critical for binding deep into to MDM2 pocket, are conserved.[7] The various contacts by MDM2 to the hydrophobic face of the peptide will shield them from the solvent and probably inhibit access to various protein-associated factors.[7]

Table 1. Sequences of wild type, single and double-stapled peptides are shown.

peptide	sequence																	
wild type	Ac	L	S	Q	E	T	**F**[a]	S	D	L	**W**	K	L	**L**	P	E	N	NH_2
cis-sah3	Ac[b]	L	S	Q	$\mathbf{X_R}$[c]	T	**F**	S	D	L	**W**	$\mathbf{X_S}$	L	**L**	P	E	N	NH_2[d]
cis-sah4	Ac	L	S	Q	E	T	**F**	$\mathbf{X_R}$	D	L	**W**	K	L	**L**	$\mathbf{X_S}$	E	N	NH_2
cis-sah8	Ac	Q	S	Q	Q	T	**F**	$\mathbf{X_R}$	N	L	**W**	R	L	**L**	$\mathbf{X_S}$	Q	N	NH_2
trans-sah8	Ac	Q	S	Q	Q	T	**F**	$\mathbf{X_R}$	N	L	**W**	R	L	**L**	$\mathbf{X_S}$	Q	N	NH_2
cis-dsah8	Ac	Q	S	Q	Q	$\mathbf{X_S}$	**F**	$\mathbf{X_R}$	N	$\mathbf{X_S}$	**W**	R	**L**	**L**	$\mathbf{X_S}$	Q	N	NH_2

a. Dark blue color denotes key hydrophobic residues Phe19, Trp23, and Leu26.

b. Ac denotes N-terminal acetylation.

c. X (red, orange, and green) denotes α -methyl-amino acids with R and S stereochemistry. (X_S, X_S) and (X_R, X_S) denotes the position of the two linkers and their respective chirality.

d. NH_2 denotes C-terminal primary amide.

4.2 Materials and Methods

Each peptide that is bound or unbound to the MDM2 protein (Appendix) was placed in an orthorhombic box and solvated with TIP3P water such that the boundary was greater than 7 Å away from the protein. An appropriate number of sodium and chlorine ions were added to maintain charge neutrality. The system was equilibrated at 300 K by a series of constrained minimizations and all atom molecular dynamics simulations using Desmond suite of programs.[11] The OPLS-2005 force field and periodic boundary conditions were employed in the simulations.[11] A 1 kcal/mol/Å^2 restraining force was applied to both the alpha carbon of residue 23 of the peptide and all alpha carbons of the protein. Sixty-four replicas[12] linearly distributed between 300 K and 600 K were employed. The replicas were evolved in parallel for 20 ns in the NPT ensemble. Replicas were exchanged every 12 ps, after the first 100 ps, between each pair of nearest neighbor trajectories. The configurations were saved in 1 ps interval during 20 ns simulation. This provided an aggregate 1280 ns of sampling. To check

convergence, we compared ensemble averages of relevant quantities over the first and the last half of the production period. Further details are provided in the appendix.

4.3 Results and Discussions

To gain an understanding of the structural profile of the stapled peptides bound to MDM2, we studied their helical contents. We report the conformation population distribution of *cis*-sah3, *cis*-sah4, *cis*-sah8, and *trans*-sah8 that are bound to MDM2, as a function of the ratio of the number of helical residues to the total number of residues in Figure 1 (and Appenix). Observe that *cis*-sah4, *cis*-sah8 and *trans*-sah8 have multipeak structures at room temperature indicating the presence of a fully helical state, several partially folded states, and one unfolded state (Appendix). Each of the two *cis* stapled peptides shows one significant peak at helicity less than 40% with 1-2 H bonds. Peptide *cis*-sah4 has two additional peaks with helicity less than 55% and 25%, each with 2 H bonds. In contrast, *cis*-sah8 has an additional peak whose helicity is less than 25% but with 2 H bonds, respectively. *cis*-dsah8 exhibits three multipeak structures in the unbound state, a feature also observed for *cis*-sah4 and *cis*-sah8.

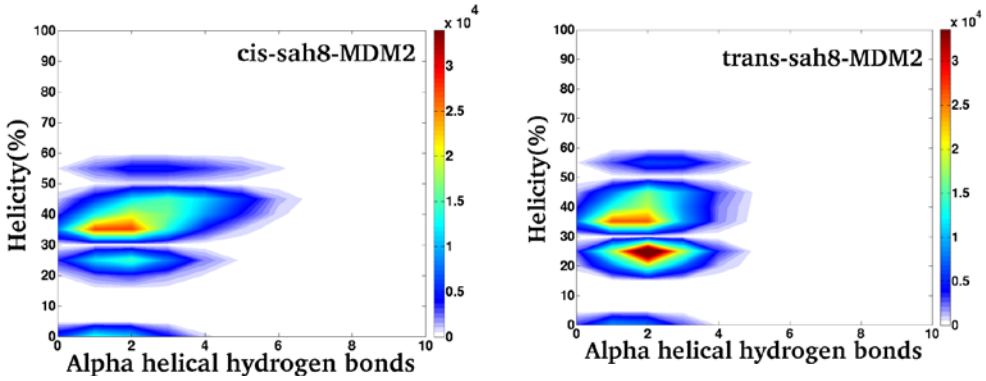


Figure 1. The contour plots of the conformational distribution of the p53 peptide analogs as a function of the number of backbone α-helical hydrogen bonds and the percent α-helical content. More red color indicates higher intensity.

We have analyzed the distribution of (ϕ,ψ) dihedral pairs of the stapled peptides bound to MDM2 (Appendix). The Ramachandran plot of *cis*-sah4 and *cis*-sah8 exhibits three peaks relative to the wild type. One of the peaks has significant population around the α-helical region. This peak is broad indicating a distorted helix. There are also two additional peaks located in the extended β-sheet region. The *cis*-dsah8 unbound has one peak in the β-sheet region, in addition to a large population in the α-helical region. The population of the extended, the unstructured or the random coil regions is significantly higher for the bound peptides than the unbound ones. We have summarized in Supplementary Material (Table 1S) the % residues whose dihedral pairs are characteristic of α-helix, i.e., within 30^o of $(-95^o, -15^o)$ and $(-35^o, -70^o)$. We find that the population of the α-helical dihedrals for the stapled peptides bound to

MDM2 is larger than that of the unbound stapled peptides.[8] For example, for *trans*-sah8 bound and unbound to MDM2, the α-helical dihedrals are 42.26% and 36.33%, respectively; corresponding values for *trans*-sah4 are 34.75% and 30.45%. In contrast, the α-helical dihedrals for *cis*-sah8 and *cis*-sah4 unbound is 32% and 40%,[8] while it is 43.80% and 41.57% in the bound state respectively. The α-helix dihedrals for *trans*-sah8 and *cis*-sah8 bound to MDM2 are comparable, while that of *cis*-dsah8 unbound is the largest. Our results indicate that there are contributions from non-α-helical structural motifs[8] to the equilibrium ensemble.

The distribution of helical occupancy across the amino acid residues is inhomogeneous (Figure 2). Sequences, over which the linker spans, tend to have the highest helical occupancy. Observe that although *trans*-sah8 is cross-linked in the same positions as *cis*-sah8 and *cis*-sah4*,* it has higher helicity content in the stapled region. The helical structures favored by the hydrocarbon cross-link are non-local. The helical content persists outside the area occupied by the cross-link, such as that observed for *cis*-sah8, *cis*-sah4, *trans*-sah8, and *cis*-dsah8 as well (Figure 2). However, the propensity to form helical structures depends on the conformation of the hydrocarbon linker, relative position of the cross-link, and specific amino acid substitutions[9,10] (Appendix).

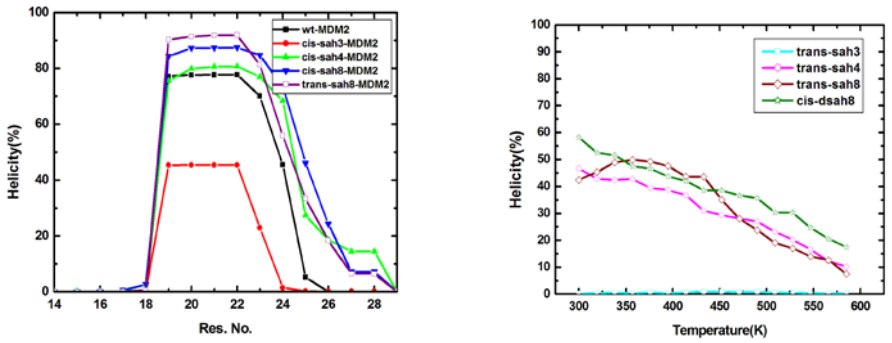


Figure 2. Comparison of percent helical occupancy versus residue number (left figure) at T = 300 K for WT, cis-sah3, cis-sah4, cis-sah8, and trans-sah8 bound to MDM2. The helix melting curves (right figure) for *trans*-sah3, *trans*-sah4, *trans*-sah8, and *cis*-dsah8 in unbound state as a function of temperature.

Analysis of structural features of the stapled peptides across a range of temperatures has been carried out to shed light, not only on the stability of the secondary structures, but also on the conformation distribution in solution. To probe this effect, we have examined the percent helical occupancy as a function of temperature (Figure 2). Wild type (WT) and *trans*-sah3 exhibit complete melting of the helix at low temperatures (Figure 2). Over a broad range of temperatures, *trans*-sah3 has a low percentage of α -helical structure, while double-stapled *cis*-dsah8 has the highest helicity at room temperatures. In some cases, such as *trans*-sah1 and *trans*-sah2, helicity is more uniformly distributed over the sequences at various temperatures (Appendix). The *trans* linker induces local helical structure that persists even as the temperature increases. Significant

helical content is observed for *cis*-dsah8 even at 575 K. The helicity of *trans*-sah8 reaches 50% around 340 K and this profile is maintained up to 400 K. A similar melting curve is observed for *trans*-sah4. The fastest decay of helical structure is *trans*-sah8 and *trans*-sah9.

We investigated the mechanism of stabilization of the stapled peptide by MDM2. An ensemble characteristic is the distribution of the fraction of native contacts Q, P(Q), which one observes during the simulation. Native contacts are formed when the distance between pairs of residues of the alpha carbon atoms are within a distance r_c taken to be 6.5 Å. The distribution P(Q) is shown in Figure 3 for WT and sah8, sah3, sah4, in which the double bond in the cross-linker is in the *cis* or *trans* conformations. When the stapled peptide is bound to MDM2, the peak in the distribution of the fraction of native contacts occurs for higher values of Q. Furthermore, the distribution is more sharply peaked for sah-8 bound to MDM2 in the *trans* conformation than when it is bound to MDM2 in the *cis* conformation. The time average value of the fraction of native contacts, $\bar{Q}$, for WT, *cis*-sah4, *cis*-sah8 and *trans*-sah8 when bound to MDM2 is 0.8743, 0.9036, 0.9178, and 0.9263 respectively. In comparison, the time average value $\bar{Q}$ for *trans*-sah4 and *trans*-sah8 in unbound state is 0.8010 and 0.7991, respectively.

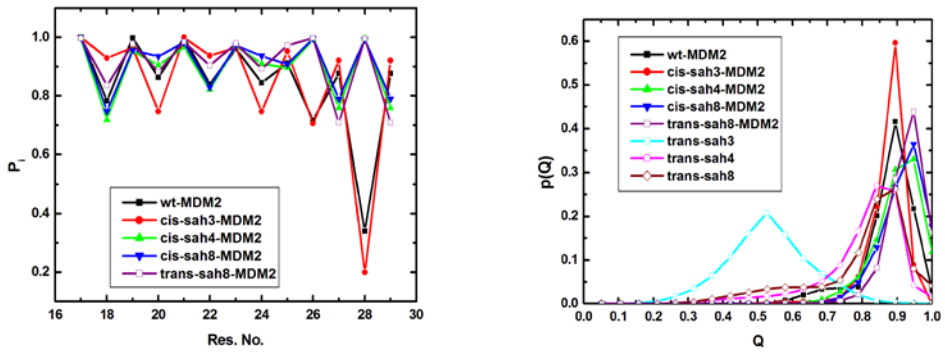


Figure 3. The average number of native contacts of the stapled peptides, P_i, versus the residue number is shown in the left. Distribution of the fraction of native contacts Q, P(Q), of the stapled peptides and compared with unbound peptides.

To elucidate the contributions to Q, we analyzed the average number of native contacts, P_i, versus residue i, for the same set of stapled peptides (Figure 3). This quantity is obtained from the contact map, cutoff r_c, and the simulation time (see Appendix). In comparison to wild type (WT), the probability to form native contacts of various residues is increased in the presence of the staple and when the stapled peptides are bound to MDM2 (Appendix). However, the stabilization effects of stapling are not local. There is an absence of a clear link between an increase or a decrease in the average number of native contacts and its arrangement in the motif.

The crystal structure of wild type p53 bound with MDM2 (PDB ID 1YCR) indicates four native contacts between the three key residues in the peptides and

MDM2; these residues are (Phe19, Gln72), (Phe19, Hie73), (Trp23, Val93), and (Leu26, Hie96). Figure 4 depicts the fraction of the native contacts observed between the three key residues Phe19, Trp23, and Leu26, and MDM2 in the course of the simulation.

A structural measure of the overall topology of the motif is given by the distribution of the end-to-end distance. Figure 4 shows the distribution of the end-to-end distance between residues 17-29 in WT-MDM2, *cis*-sah8-MDM2, *trans*-sah8-MDM2, and *trans*-sah8 unbound. Observe that the various distributions are bimodal. One of the peaks in the distribution of the stapled peptides, bound or unbound, is centered around 21 Å, a value that is localized around the native value of 23.10 Å obtained from crystal structure (PDB code 1YCR) of MDM2 bound to p53.

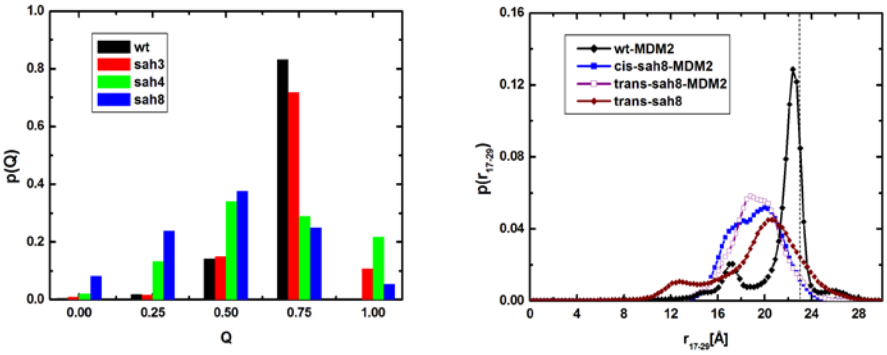


Figure 4. The distribution (left figure) of the fraction of native contacts Q, P(Q), of the three key residues of the peptide and MDM2. The right figure indicates the end-to-end distribution of p53 residues 17-29 bound to MDM2 and *trans*-sah8

unbound. The vertical dash line indicates the value (23.10 Å) of the end-to-end distribution obtained from the crystal structure of p53-MDM2

According to the analysis of experimental[4] and computational result (Appendix), it is clearly showed that the stapled peptide 4 shows the best binding affinity with protein MDM2, stapled peptide 8 shows higher binding affinity than wild type p53 and stapled peptide 3, but lower than stapled peptide 4. The main difference of the binding free energies arise from the entropic part; this is consisted with the presumption that the hydrocarbon linker will pre-organize and stabilize the structure of p53 peptide which in turn will reduce entropy changes during binding with protein MDM2. Also observe that *cis* cross-linked peptide shows better binding affinity then *trans* cross-linked peptide.

4.4 Summaries

In summary, we have investigated α-helical conformation and structural stability of single and double stapled all-hydrocarbon cross-linked p53 peptides in solution and when bound to MDM2. These peptides allowed us to quantitatively unravel the effects of the peptide sequence, the stereochemistry of the cross-linker, the conformation of the double bond in the alkene bridge, and the length of the tether, on the relative stability of the α-helix structure. In future work, we will explore various dynamical variables that could be utilized as probes for the mechanism of stabilization induced by the hydrocarbon linker. We will, for

example, use the autocorrelation function of the end-to-end distance to investigate whether the staple acts as a frequency sieve when bound to MDM2.

References

1. Joerger, A.C.; Ferscht, A.R. “The tumor suppressor p53: from structures to drug discovery.” in *Cold Spring Harbor Perspectives in Biology*. Levine, A.J and Lane, D. eds. (Cold Spring Harbor Laboratory Press, **2010**; pages 1-20).
2. Walensky, L.D.; Verdine, G.L., *Clin. Cancer Res*. **2007**, 13, 7264-7270.
3. Onuchic, J.N.; Luthey-Schulten, Z.; Wolynes, P.G., *Ann. Rev. Phys. Chem*. **1997**, 48, 545-600.
4. Bernal, F.; Tyler, A.F.; Korsmeyer, S.J.; Walensky, L.D.; Verdine, G.L., *J. Am. Chem. Soc.* **2007**, 129, 2456-2457.
5. Schafmeister, E.; Po, J.; Verdine, G.L., *J. Am. Chem. Soc*. **2000**, 122, 5891-5892.
6. Kritzer, J.A.; Lear, J.D.; Hodsdon, M.E.; Schepartz, A., *J. Am. Chem. Soc*. **2004**, 126, 9468-9469.
7. Kussie, P.H.; Gorina, S.; Marechal, V.; Elenbaas.; Moreau, J.; Levine, A.J.; Pavletich, N.P., *Science* **1996**, 274, 948-953.
8. Guo, Z.; Mohanty, U.; Noehre, J.; Sawyer, T.; Sherman, W.; Krolov, G., *Chem. Biol. Drug. Des*. **2010**, 75, 348-359; see various references cited therein.
9. Moretto, A.; Crisma, M.; Kaptein, B.; Broxterman, Q.B.; Toniolo, C., *Biopolymers* **2006**, 84, 553-565.
10. De Poli, M.; Moretto, A.; Crisma, M.; Peggion, C.; Formaggio, F.; Kaptein, B.; Broxterman, Q.B.; Toniolo, C., *Chemistry* **2009**, 15, 8015-8025.
11. Desmond v2.2, Schrödinger, Inc.: Portland, OR.

12. Sugita, Y.; Okamoto, Y., *Chem. Phys. Lett*. **1999**, 314, 141-151.
13. Postma, J.; Berendsen, H.; Haak, J., *Faraday Symp. Chem. Soc*. **1982**, 17, 55-67.
14. Miyamoto, S.; Kollman, P.A., *PROTEINS: Structure, Function, and Genetics* **1993**, 16:226-245.
15. Massova, I.; Kollman, P.A., *Perspectives in Drug Discovery and Design*, **2000**, 18, 113-135.
16. Onufriev, A.; Case, D.A.; Bashford, D., *J. Comp. Chem.,* **2002**, 23, 14, 1297-1304.
17. Kaminski, G.A.; Friesner, R.A.; Tirado-Rives, J.; Jorgensen, W.L., *J. Phys. Chem. B* **2001**, 105, 6474-6487.
18. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G., *J. Chem. Phys*. **1995**, 103: 8577-8593.
19. Martyna, G.J.; Tobias, D.J.; Klein, M.L., *J. Chem. Phys.* **1994**, 101: 4177–4189.
20. Hoover, W.G., *Phys. Rev. A* **1985**, 31, 1695-1697.
21. Hamacher, K.; Hubsch, A.; McCammon, J. A., *J. Chem. Phys*. **2006**, 214, 164907164914.

Appendix A

Free energy perturbation (FEP)

Free energy perturbation (FEP) approach[13] was used to calculate the absolute binding free energies between the p53 peptides and MDM2. 10.0 Å water buffer was set in the simulation box. 2.0 ns production run was carried out in the NPT ensemble. Further details of the simulations are as described in the chapter.

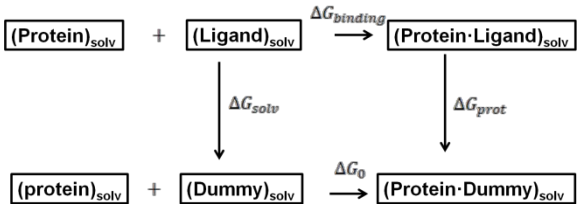


Figure A1. The absolute binding free energy of p53 stapled peptide binding to MDM2 is determined via the thermodynamic cycle. Figure is adopted from reference 14.

The total binding free energies is obtained via the thermodynamic cycle (Desmond v2.2)[11] as shown in Figure A1. ΔG_{solv} denotes the free energy for mutating the ligand to a reference molecule in solution, which is an estimate of the solvation free energy of respective part. ΔG_{prot} is the free energy for mutating the ligand to a reference molecule in the complex.[14] The reference molecule has all interactions of the original molecules except van der Waals or electrostatic interaction (the thermodynamic integration parameter λ=0).[14] Since

there are no interactions, ΔG_o is zero. The absolute binding free energy is then obtained from the relation[14]

$$\Delta G_{binding} = \Delta G_{solv} - \Delta G_{prot}$$

Molecular mechanics (MM)-GBSA

The absolute binding free energies between the various stapled p53 peptides and MDM2 was also determined in the MM-GBSA approximation[15] which is the most commonly used method to calculate the protein-protein interactions. In MM-GBSA, molecular mechanism calculation (MM) is combined with the Generalized Born model (GB) and the hydrophobic solvent accessible surface area term (SA).

The GB model is an approximation to the Poisson equation, the later is computationally less than convenient. In a protein system, the GBSA describes a molecule as a set of spheres with different dielectric constants between internal solvent and solvent external to the sphere. The GB model[15,16] the electrostatic free energy ΔG_{el} can be approximated by:

$$\Delta G_{el} \approx \Delta G_{GB} = -\sum_i \frac{q_i^2}{2R_i}\left(1-\frac{1}{\varepsilon_w}\right) - \frac{1}{2}\sum_{i,j,i\neq j}^{N} \frac{q_i q_j}{f^{GB}(r_{ij},R_i,R_j)}\left(1-\frac{1}{\varepsilon_w}\right)$$

where $f^{GB} = \sqrt{r_{ij}^2 + R_{ij}^2 e^{-D}}$, $D = \left(\frac{r_{ij}}{2R_{ij}}\right)^2$, $R_{ij} = \sqrt{R_i R_j}$. ε_W is the dielectric constant of the solvent. r_{ij} is the distance between two charges i and j, and R_i is the effective Born radii. The effective Born radii of an atom characterizes its

degree to which it is buried inside the solute. An accurate effective Born radii estimation is a critical input for the GB model.[15,16]

During the conformational sampling, 10.0 Å water buffer was set in simulation box. A 40.0 ns molecular dynamics simulation (Desmond v2.2)[11] was carried out in the NVT ensemble. The entropic part of free energy was calculated in the rigid rotor harmonic oscillator approximation (RRHO).[14,15] The absolute binding free energy is obtained from the relation[14,15]

$$\Delta G_{binding} = \Delta E - T\Delta S$$

Binding free energy of p53 peptides to MDM2

In Table 1A we have summarized the absolute binding free energies of p53 peptides to MDM2 obtained by free energy perturbation (FEP), MM-GBSA, and compared it with experimental data from Bernal *et al*.[4], respectively.

The absolute binding free energy is derived from the differences of solvation free energy of the ligand peptides when the receptor MDM2 is present or not present. These calculations (Table 2A) show that the solvation free energy of stapled peptide 8 is noticeably lower than other peptides; this is consistent with the charges these peptides carry. Stapled peptide 8 has a charge of +1, while other peptides have a charge of -2. As a result, one expects that stapled peptide 8 is easier to solvate in an aqueous solution. From the FEP results (second column), stapled peptide 4 shows the highest binding affinities, and stapled peptide 3 shows the lowest binding affinity. Stapled peptide 8 is higher than wild type and stapled peptide 3, but is lower than stapled peptide 4. The trends of the

predictions based on free energy perturbation are consistent with the experimental data; however, the values are about 30 kcal/mol higher in each case, which is also found using the MM-GBSA approximation.

In MM-GBSA and rigid rotor harmonic oscillator approximation[15,16,17], the binding affinities were divided into enthalpic and entropic parts as shown in Table 3A. The results of MM-GBSA for enthalpy and rigid rotor harmonic oscillator approximation for entropy (Table 3A) are consistent with the free energy perturbation calculations and the trends in the experimental data. The stapled peptide 4 shows the highest binding affinity, while stapled peptide 8 shows higher binding affinity than wild type p53 and stapled peptide 3, but lower than stapled peptide 4. The main difference of the binding free energies arise from the entropic part; this is consisted with the assumption that the hydrocarbon linker will pre-organize and stabilize the structure of p53 peptide; this will reduce entropy changes during binding with protein MDM2. Finally, we observed that the *cis* cross-linked peptide shows better binding affinity then *trans* cross-linked peptide.

Table 1A. Absolute binding free energies of various stapled p53 peptides binding to MDM2 are summarized. Experimental value from reference 4.

Peptides	$\Delta G_{binding}$ (FEP) (kcal/mol)	$\Delta G_{binding}$ (MM/GBSA) (kcal/mol)	$K_d(nM)$ (experimental) (kcal/mol)
wt-MDM2	-51.44	-62.74	410±19 (-20.00)
cis-sah3-MDM2	-45.14	-63.65	1200±89 (-18.54)
cis-sah4-MDM2	-62.88	-71.72	0.92±0.11 (-28.29)
cis-sah8-MDM2	-52.74	-64.70	55±11 (-22.73)
trans-sah4-MDM2	-52.40	-58.01	N/A
trans-sah8-MDM2	-40.16	-60.11	N/A

*The experimental helicities of p53 peptides were obtained through Circular Dichroism spectroscopy data by Verdine *et al.* In which, compounds were dissolved in H_2O, and the concentrations of compounds ranged from 10-50 μM at 20º C.[4]

Table 2A. Absolute binding free energies of various stapled peptides based on free energy perturbation are compared with experimental data.

Peptides	ΔG_{solv} (kcal/mol)	ΔG_{prot} (kcal/mol)	$\Delta G_{binding}$ (kcal/mol)	$K_d(nM)$ Bernal *et al*[4] (kcal/mol)
wt-MDM2	213.46(4.7)	264.90(1.87)	-51.44	410±19 (-20.00)
cis-sah3-MDM2	160.20(2.6)	205.34(2.09)	-45.14	1200±89 (-18.54)
cis-sah4-MDM2	208.24(4.5)	271.12(1.90)	-62.88	0.92±0.11 (-28.29)
cis-sah8-MDM2	91.47(3.8)	144.21(1.63)	-52.74	55±11 (-22.73)
trans-sah4-MDM2	222.97(1.47)	275.37(1.16)	-52.40	N/A
trans-sah8-MDM2	106.79(2.02)	146.95(1.22)	-40.16	N/A

*The experimental helicities of p53 peptides were obtained through Circular Dichroism spectroscopy data by Verdine *et al.* In which, compounds were dissolved in H_2O, and the concentrations of compounds ranged from 10-50 μM at 20º C.[4]

Table 3A. Absolute binding free energies of various stapled peptides using MM/GBSA approximation are compared with experimental data.

Peptides	ΔE (kcal/mol)	$-T\Delta S$ (kcal/mol)	$\Delta G_{binding}$ (kcal/mol)	$K_d(nM)$ Bernal *et al*[4] (kcal/mol)
wt-MDM2	-137.29(14.70)	74.55	-62.74	410±19 (-20.00)
cis-sah3-MDM2	-137.19(10.36)	73.54	-63.65	1200±89 (-18.54)
cis-sah4-MDM2	-135.99(12.43)	64.27	-71.72	0.92±0.11 (-28.29)
cis-sah8-MDM2	-130.30(11.15)	65.60	-64.70	55±11 (-22.73)
trans-sah4-MDM2	-121.60(13.62)	63.59	-58.01	N/A
trans-sah8-MDM2	-130.82(12.50)	70.71	-60.11	N/A

*The experimental helicities of p53 peptides were obtained through Circular Dichroism spectroscopy data by Verdine *et al.* In which, compounds were dissolved in H_2O, and the concentrations of compounds ranged from 10-50 μM at 20º C.[4]

Appendix B

B1. Double staple

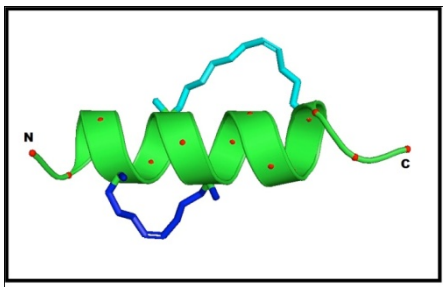


Figure B1. Double *cis* hydrocarbon linkers (blue color) is stapled to p53 peptide (green color). The alpha carbons are shown in red.

B2. Materials and Methods.

The sequences of a 16-residue transactivation domain of p53 peptides used in this study are shown in Table 1. The hydrocarbon tether stapled to the peptide at either i,i+7 or i,i+4 residues was incorporated, with the double bond placed in a *cis* or *trans* configuration, using Maestro suite of programs.[12] Since there are no crystal structures of stapled peptides bound to MDM2, we used the crystal structure of the wild type (WT) p53 bound to MDM2 (PDB: 1YCR) as our initial structure for replica exchange molecular dynamic simulations.[12] The stapled peptide bound to MDM2 complex was created by mutating the crystal structure of the wild type p53 peptide bound to MDM2 (PDB code 1YCR). Protonation states was assigned based on pH=7. Acetyl and carboxamide moieties were included at the N-and C-termini of the p53 peptides as well as the MDM2 protein chain.[8] The bound complex was placed in an orthorhombic cell

with TIP3P water. Sodium and chlorine ions was added so that one has charge neutrality.[8] OPLS 2005 force field was employed to model the peptides and water interactions.[17] The system was equilibrated with joint energy minimization and restrained MD simulations.[11] Long-range electrostatic interactions were calculated by particle-mesh Ewald method.[18] Short range electrostatic and Van der Waals interactions were truncated at 7.0Å. To control the pressure, Martina–Tobias–Klein method was used.[19] Constant simulation temperature was maintained by Nose–Hoover thermostats.[20]

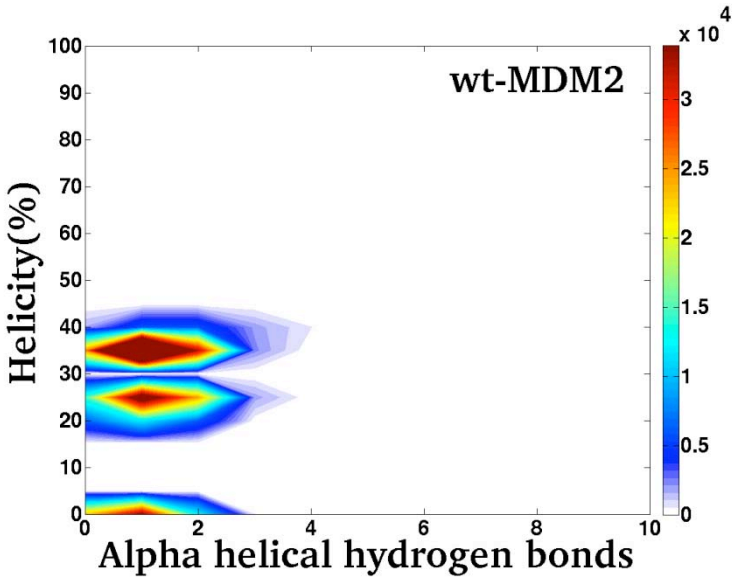
In order to keep the stapled peptide bound to MDM2 in the hydrophobic pocket, 1 kcal/mol/Å^2 restraining force was applied to both the alpha carbon of residue 23 of the peptide and all alpha carbons of the protein. By restraining the protein, we did not need to use too many replicas in the simulation, as effectively the system would have fewer mobile degrees of freedom. We employed 64 replicas of the system. The replicas[12] were evolved in parallel for 20 ns in the NPT ensemble with linear temperature spacing between 300 K and 600 K. After the first 100 ps, we carried out replica exchange every 12 ps between pairs of adjacent trajectories so as to equilibrate the system. After 10 ns period, we saved configurations every 1-ps time interval for the final 10 ns run.

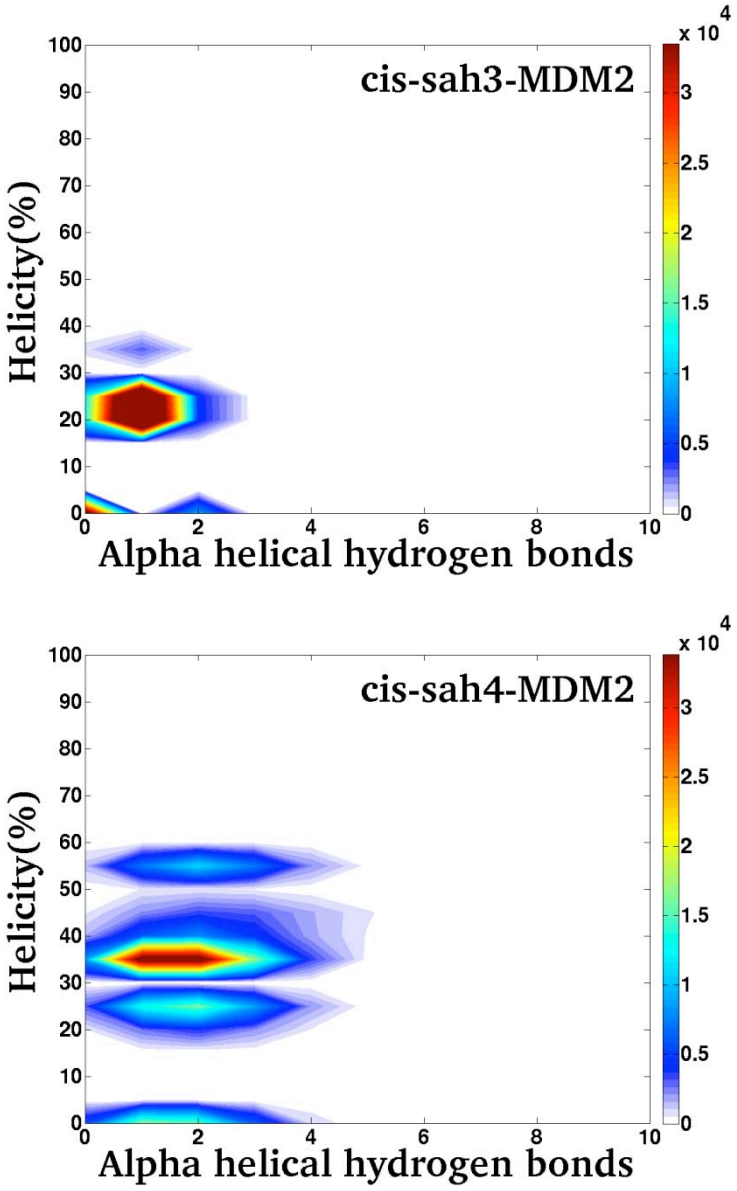
Native contacts are formed when the distance between pairs of residues of the alpha carbon atoms are within a distance r_c taken to be 6.5 Å. The average number of native contacts, P_i, was computed from the simulation using the relation[21]

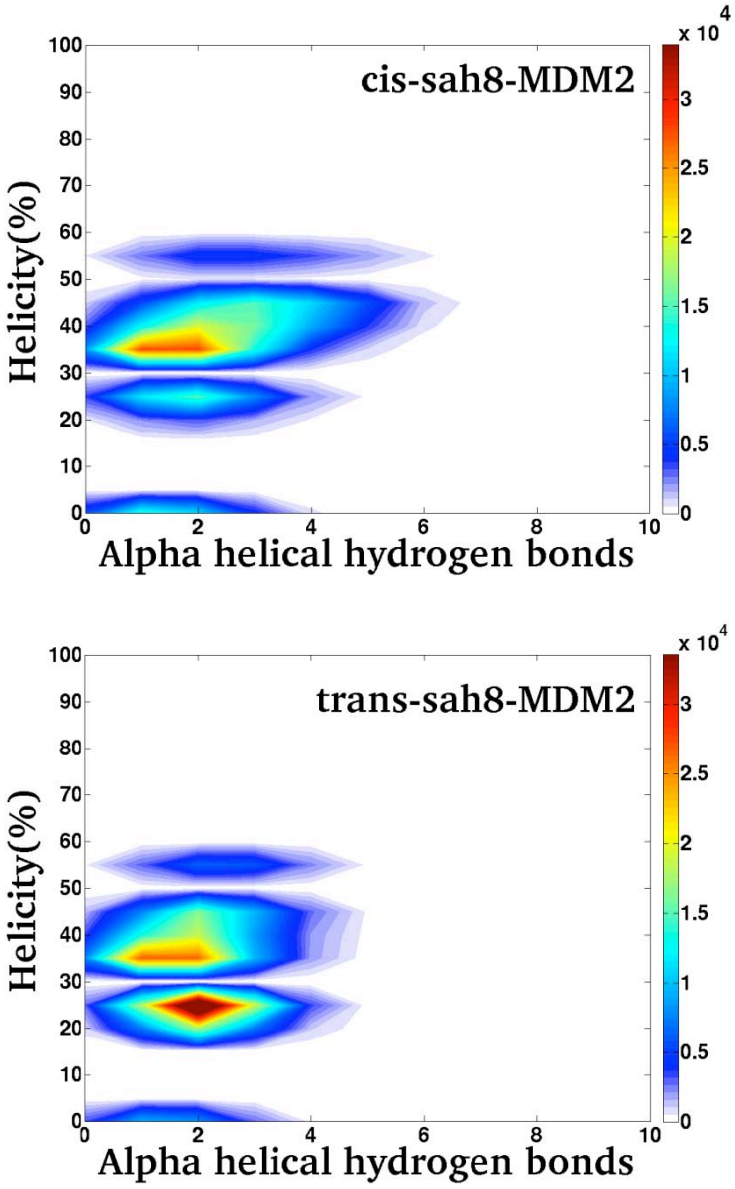
$$P_i = \sum_i \sum_j \Delta_{ij} \Theta(x) / t_o \sum_j \Delta_{ij}, \quad (1)$$

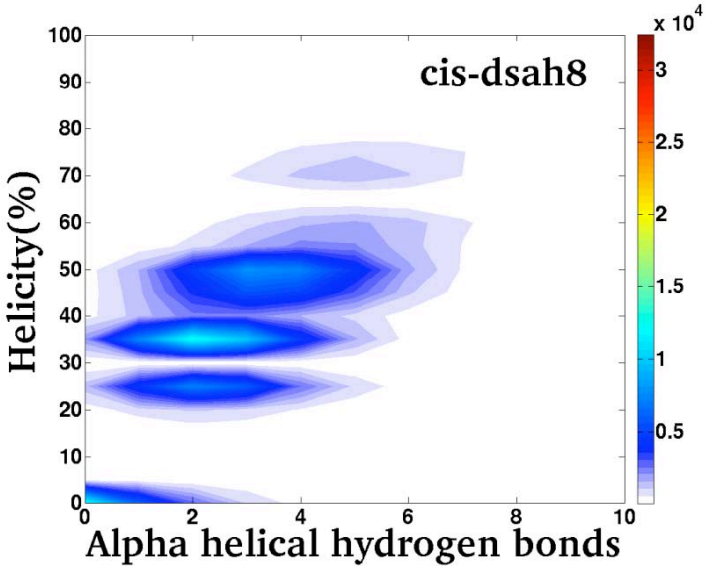
where the contact map of residues i and j is denoted by Δ_{ij}, $\Theta(x)$ is the characteristic function, t_o is the simulation steps, and $x = r_c - r_{ij}(t)$.

B3. Conformation population plots. The contour plots of the conformational distribution of the p53 peptide analogs as a function of the number of backbone α-helical hydrogen bonds and the percent α-helical content, for WT and various *cis* and *trans* single and double stapled peptides bound to MDM2. Higher color indicates higher intensity.

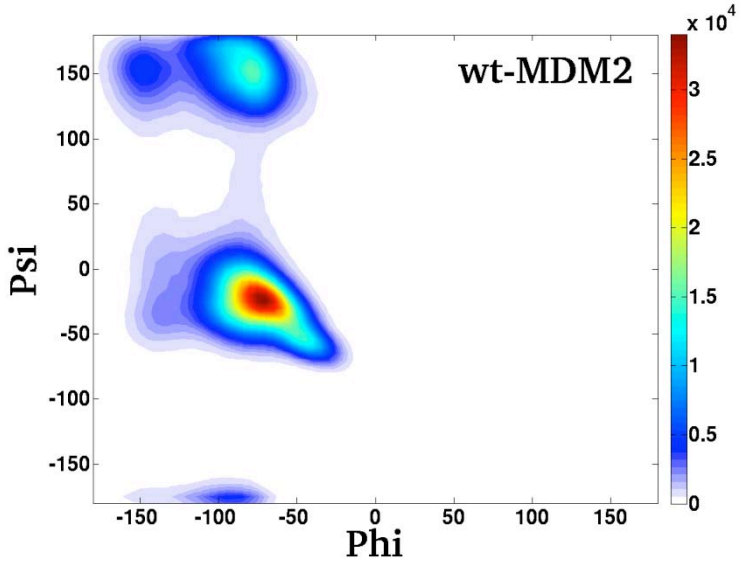


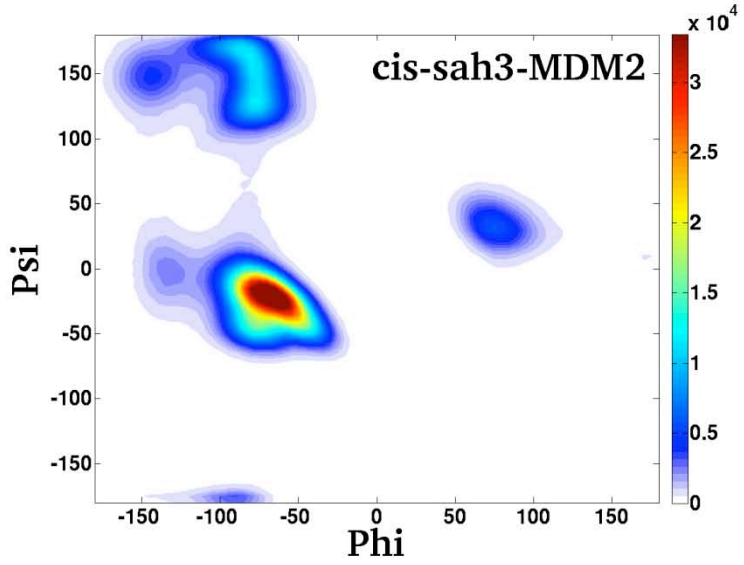

cis-sah3-MDM2
Helicity(%)
Alpha helical hydrogen bonds
x 10^4
cis-sah4-MDM2
Helicity(%)
Alpha helical hydrogen bonds
x 10^4

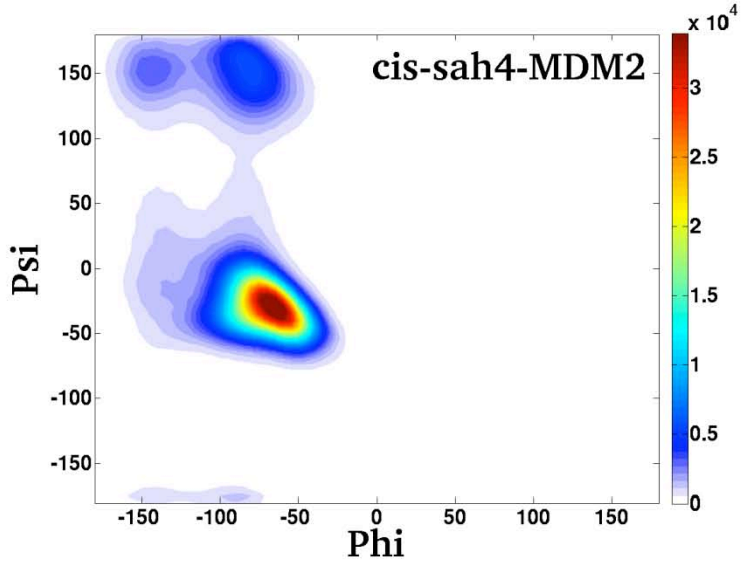

cis-sah8-MDM2
trans-sah8-MDM2
Helicity(%)
Alpha helical hydrogen bonds
x 10^4
0
0.5
1
1.5
2
2.5
3
0
2
4
6
8
10
0
10
20
30
40
50
60
70
80
90
100

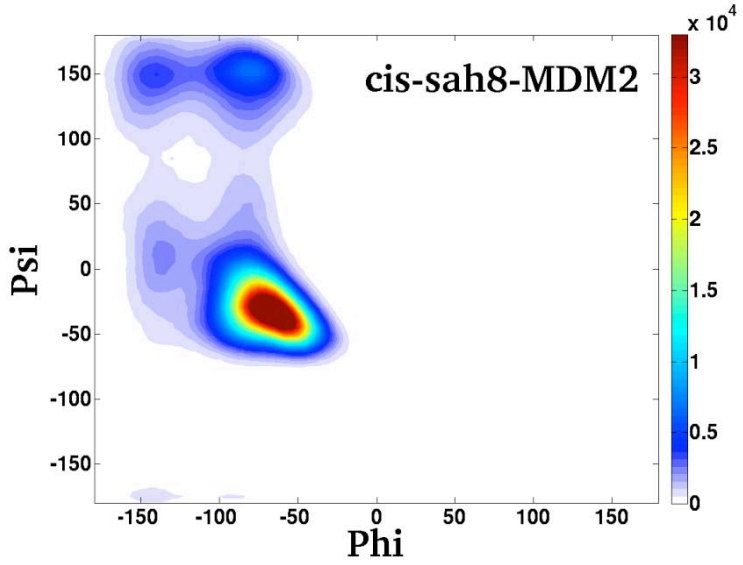

cis-dsah8
Helicity(%)
Alpha helical hydrogen bonds
x 10^4
100
90
80
70
60
50
40
30
20
10
0
0
2
4
6
8
10
3
2.5
2
1.5
1
0.5

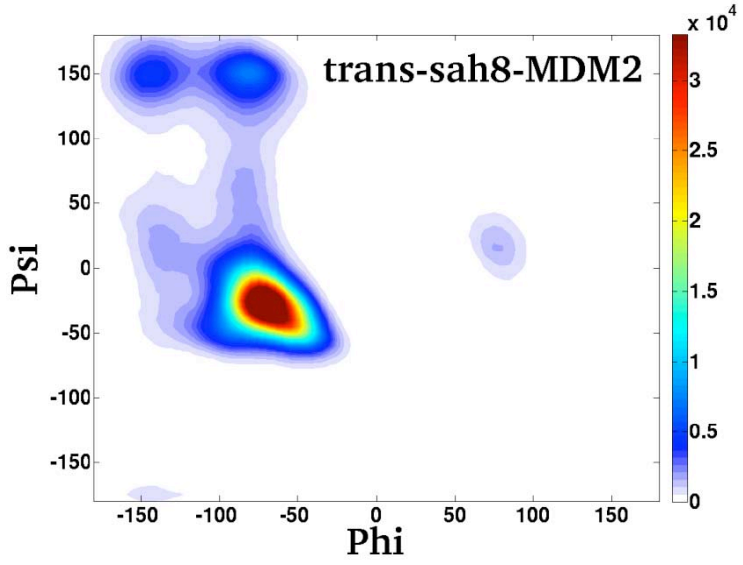
B4. Ramachandran plots. The Ramachandran plot of the distribution of the backbone dihedrals for WT and *cis* and *trans* single and double stapled peptides bound to MDM2. More red color indicates higher intensity.

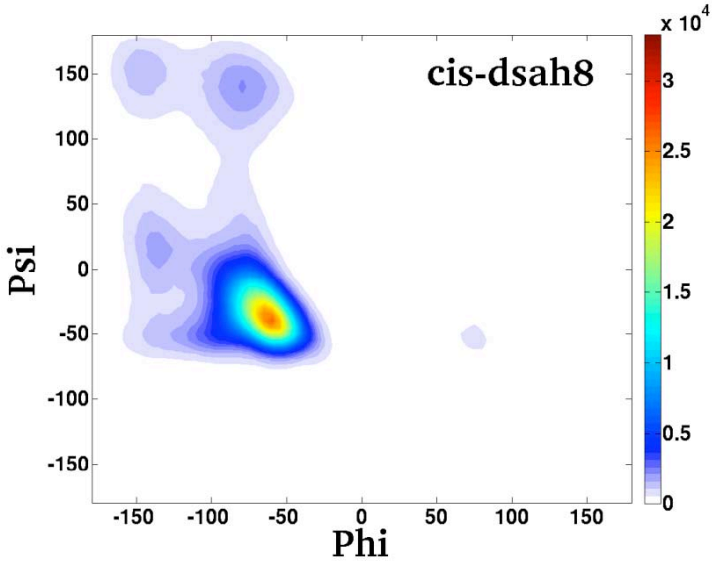



cis-sah3-MDM2
x 10^4
3
2.5
2
1.5
1
0.5
0
150
100
50
0
-50
-100
-150
Psi
-150
-100
-50
0
50
100
150
Phi

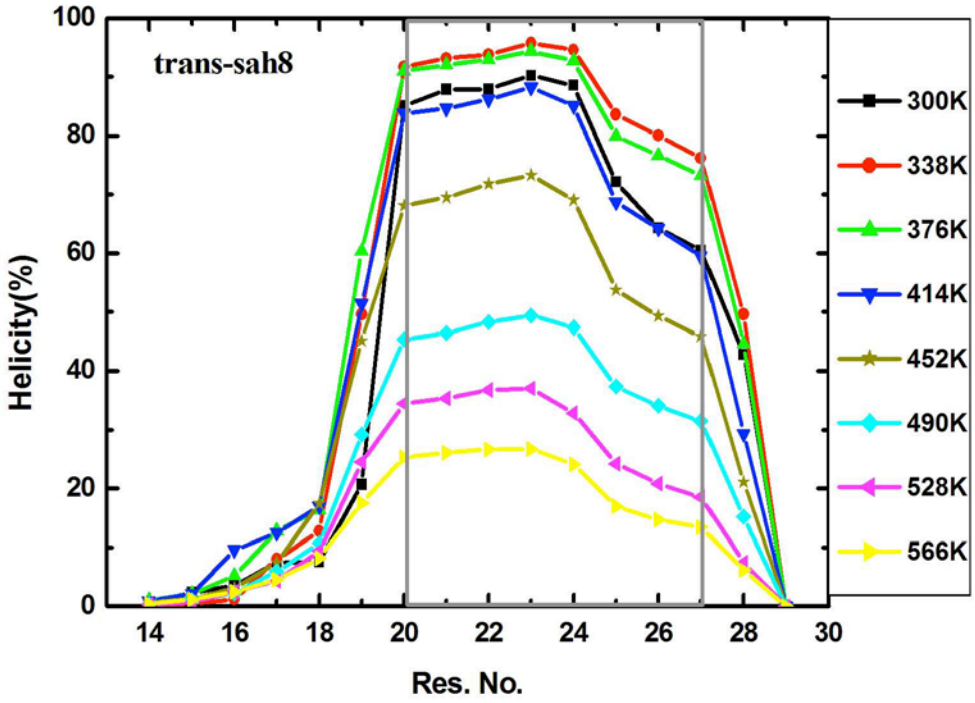

cis-sah4-MDM2
x 10^4
3
2.5
2
1.5
1
0.5
0
150
100
50
0
-50
-100
-150
Psi
-150
-100
-50
0
50
100
150
Phi

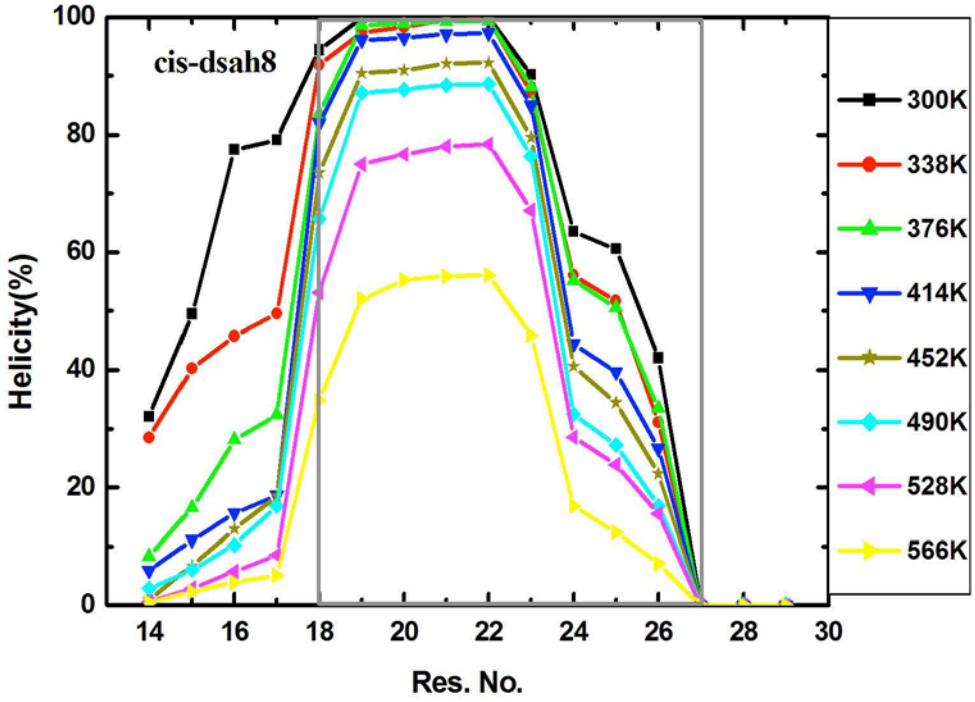

cis-sah8-MDM2
Psi
Phi
x 10^4
3
2.5
2
1.5
1
0.5
0
150
100
50
0
-50
-100
-150


trans-sah8-MDM2
Psi
Phi
x 10^4
3
2.5
2
1.5
1
0.5
0
150
100
50
0
-50
-100
-150


cis-dsah8
x 10^4
3
2.5
2
1.5
1
0.5
0
150
100
50
0
-50
-100
-150
Psi
-150
-100
-50
0
50
100
150
Phi

B5. Percentage helicity versus residue number of *trans*-sah8 and *cis*-dsah8 at various temperatures.




cis-dsah8
Helicity(%)
Res. No.
300K
338K
376K
414K
452K
490K
528K
566K

B6. Average number of native contacts, P_i, versus residue i, for *trans*-sah8 (unbound) and *cis*-dsah8 (unbound).

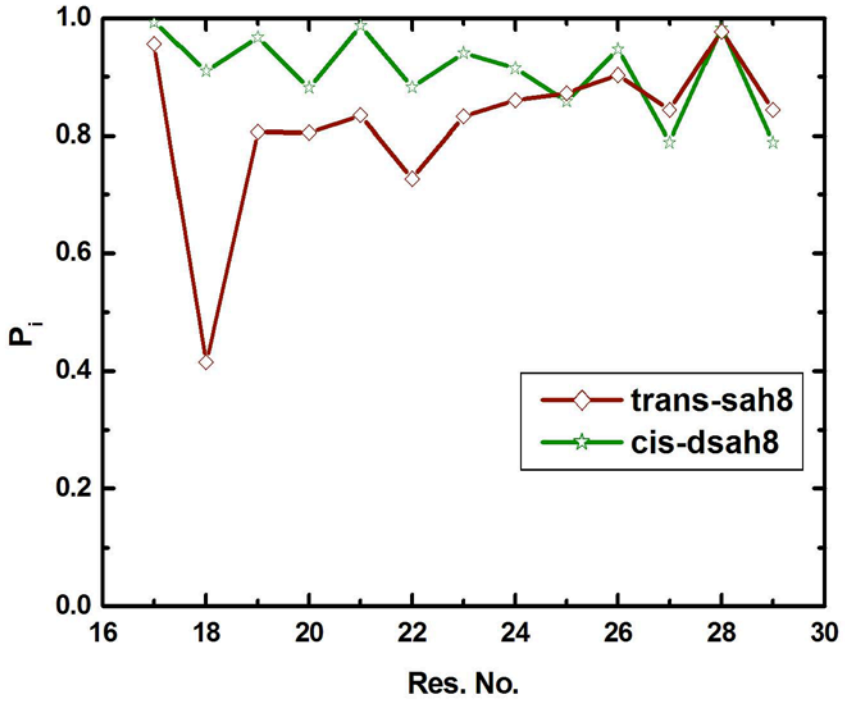


Table 1B. α -helix dihedrals and backbone H-bonds for single and double stapled peptides.

peptide	α -helix dihedrals	backbone H-bonds
wild type	30.19%	3.38
cissah3	33.13%	3.54
cissah4	41.57%	4.45
cissah8	43.80%	4.63
transsah8	42.26%	4.64
cisdsah8	44.99%a	5.57a

a. The data for *cis*-dah8 is for unbound state.